Lecture notes for Math33B: Differential Equations
Last revised June 1, 2020

Allen Gehret

Mikhail Hlushchanka

Author address:

Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095
E-mail address: allen@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095
E-mail address: m.hlushchanka@math.ucla.edu

Contents

Abstract

The objective of this class is to experience an introduction to the rich, complex, and powerful subject of *Ordinary Differential Equations (ODEs)*. Specifically:

(1) Develop a working familiarity with linear algebra to the extent we need it for the differential equations we shall consider. Linear algebra serves us as a very robust backend for handling all higher-dimensional linear issues which will arise.
(2) Learn how to solve a reasonably large class of differential equations. Most differential equations cannot be solved (the solutions can only be approximated with computers, which is a story for a different math class), but we will teach you many of the differential equations for which we can find exact solutions.
(3) Observe and investigate real-world applications which are governed by differential equations.
(4) Study qualitative properties of both the differential equations we can solve and those we cannot.

The textbook for the course is *Differential Equations* Second Edition, by John Polking, Albert Boggess, and David Arnold [**2**]. These notes are based on this textbook, except for the sake of time we only include a select curated portion of the textbook material in these notes. Any and all comments, typos, errors, questions, suggestions are enthusiastically welcome!

Last revised June 1, 2020.
2010 *Mathematics Subject Classification.* Primary .
The first author is supported by the National Science Foundation under Award No. 1703709.

List of Figures

Introduction

The prerequisite for this course is *Math31B: Integration and Infinite Series*. Consequently, we will assume you have a working familiarity with the basic properties of differentiation and integration of common elementary functions (although we will review the tools which are most relevant for us). In this class we will put these existing tools to work to help us solve so-called *differential equations*. We begin with a simple example of a differential equation:

Question 0.0.1. *Find a differentiable function $y : \mathbb{R} \to \mathbb{R}$ which satisfies the following:*

(1) $y'(t) = \exp(t)$ *for all* $t \in \mathbb{R}$*, and*
(2) $y(0) = 10$.

Answer. From (1) we know that the function $y(t)$ must be of the form $y(t) = \exp(t) + C$ for some fixed $C \in \mathbb{R}$. By (2) we know that $y(0) = \exp(0) + C = 1 + C = 10$. Thus $C = 9$ and so $y(t) = \exp(t) + 9$. □

Question 0.0.1 illustrates a paradigm for differential equations in general. Namely, we will often be given the following information:

(1) Information about an unknown function y's derivative (or second derivative, etc.), for instance, saying "$y'(t) = \exp(t)$"
(2) Information about specific function values of y (or y', y'', etc.), for instance, saying "$y(0) = 10$".

Then the game will then be to use this information to determine the unknown function y as specifically as we can. Before we go any further, we make the following declaration:

You will not be able to solve most differential equations.

This is by no means a commentary on anyone's mathematical abilities, we simply want to bring you up to speed with a cold hard fact of life: *most differential equations are impossible (for anyone) to solve exactly.* However, we will study in detail many simple differential equations which we can solve exactly. Fortunately, the differential equations we will study also have many practical real-world applications.

What about the non-solvable differential equations? Not all hope is lost in this case. Indeed, for practical real-world applications you generally only need a sufficiently accurate approximation of a solution. Luckily this is something that computers are very good at and this is a very active area of applied mathematics. We will not go down this rabbit-hole in this class, but it helps to be aware of this remedy so you are not too discouraged if and when you encounter an impossible differential equation.

Algebraic equations

In this section we will review the state of affairs for one-variable algebraic equations. Recall that a one-variable algebraic equation is an equation of the form:

$$p(X) \;=\; 0,$$

where p is a polynomial and X is a variable. A **solution** to this equation is a specific real number $x \in \mathbb{R}$ which has the property that $p(x) = 0$ (i.e., when we plug in the number x into p, it evaluates to the number 0).

We also hope to make a general point in this section: that even for algebraic equations (i.e., a differential equation with *no* derivatives), things become very complicated and eventually impossible very quickly.

Linear equations. A **linear equation** (in one variable) is an equation of the form:

$$a_1 X + a_0 \;=\; 0 \quad (\text{where } a_1, a_0 \in \mathbb{R})$$

If $a_1 \neq 0$, then this has exactly one solution, namely:

$$x \;:=\; -\frac{a_0}{a_1}.$$

If $a_1 = 0$, then this has either zero solutions (for instance, if $a_0 \neq 0$), or infinitely many solutions (for instance, if $a_0 = 0$ then every $x \in \mathbb{R}$ is a solution). These observations foreshadow various features of systems of linear equations in multiple variables which we will study in Chapter 1.

Quadratic equations. A **quadratic equation** is an equation of the form:

$$a_2 X^2 + a_1 X + a_0 \;=\; \qquad (\text{where } a_2, a_1, a_0 \in \mathbb{R})$$

If $a_2 \neq 0$, then the **quadratic formula** yields solutions:

$$x_1 \;:=\; \frac{-a_1 + \sqrt{a_1^2 - 4a_2a_0}}{2a_2} \quad \text{and} \quad x_2 \;:=\; \frac{-a_1 - \sqrt{a_1^2 - 4a_2a_0}}{2a_2}$$

Recall that three things can happen depending on the sign of the **discriminant** $a_1^2 - 4a_2a_0$:

(Case 1) If $a_1^2 - 4a_2a_0 > 0$, then $x_1 \neq x_2$ are two *real* solutions.
(Case 2) If $a_1^2 - 4a_2a_0 = 0$, then $x_1 = x_2$ is a single real solution (of multiplicity two).
(Case 3) If $a_1^2 - 4a_2a_0 < 0$, then $x_1 \neq x_2$ are two distinct solutions, however, they will be complex solutions and not real solutions.

You are expected to be able to use the quadratic formula to solve quadratic equations in this class.

Cubic equations. A **cubic equation** is an equation of the form:

$$a_3 X^3 + a_2 X^2 + a_1 X + a_0 \;=\; 0 \quad (\text{where } a_3, a_2, a_1, a_0 \in \mathbb{R})$$

You were probably never taught the formula for the cubic equation in school. This is for good reason: it's complicated! You do not need it for this class either, but in case you are curious, here it is: if $a_3 \neq 0$, then the three solutions are

$$x_k \;=\; -\frac{1}{3a_3}\left(a_2 + \xi^k C + \frac{\Delta_0}{\xi^k C}\right), \quad \text{for } k = 0, 1, 2$$

where

$$\begin{aligned} \xi &:= \frac{-1+\sqrt{-3}}{2} \\ \Delta_0 &:= a_2^2 - 3a_3a_1 \\ \Delta_1 &:= 2a_2^3 - 9a_3a_2a_1 + 27a_3^2a_4 \\ C &:= \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\ &\quad \text{(choose either + or − provided } C \neq 0) \end{aligned}$$

Here there can either be three, two, or one distinct solution, and the solutions can be either real or complex, much like the quadratic equation.

Quartic equations. A **quartic equation** is an equation of the form:

$$a_4X^4 + a_3X^3 + a_2X^2 + a_1X + a_0 = 0 \quad (\text{where } a_4, a_3, a_2, a_1, a_0 \in \mathbb{R})$$

The general solution for the quartic equation is even more complicated than the equation for the cubic. You definitely do not need to know it, but in case you are curious here it is: if $a_4 \neq 0$, then the four solutions are:

$$\begin{aligned} x_{1,2} &:= -\frac{a_3}{4a_4} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{q}{S}} \\ x_{3,4} &:= -\frac{a_3}{4a_4} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{q}{S}} \end{aligned}$$

where

$$\begin{aligned} p &:= \frac{8a_4a_2 - 3a_3^2}{8a_4^2} \\ q &:= \frac{a_3^3 - 4a_4a_3a_2 + 8a_4^2a_1}{8a_4^3} \\ S &:= \frac{1}{2}\sqrt{-\frac{2}{3}p + \frac{1}{3a}\left(Q + \frac{\Delta_0}{Q}\right)} \\ Q &:= \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\ \Delta_0 &:= a_2^2 - 3a_3a_1 + 12a_4a_0 \\ \Delta_1 &:= 2a_2^3 - 9a_3a_2a_1 + 27a_3^2a_0 + 27a_4a_1^2 - 72a_4a_2a_0 \\ &\quad \text{(with special cases if } S = 0 \text{ or } Q = 0) \end{aligned}$$

Quintic (and higher degree) equations. A **quintic equation** is an equation of the form:

$$a_5X^5 + a_4X^4 + a_3X^3 + a_2X^2 + a_1X + a_0 = 0 \quad (\text{where } a_5, a_4, a_3, a_2, a_1, a_0 \in \mathbb{R})$$

You might be expecting an even longer and more complicated formula for the five solutions to a quintic equation, but actually it is known that this is impossible. In fact, there is a theorem which tells us that this is impossible:

Theorem 0.0.2 (Galois)**.** *Suppose $n \geq 5$. Then there is no general formula using radicals ($\sqrt{}$, $\sqrt[3]{}$, $\sqrt[4]{}$, ...) which gives the solutions to*

$$a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \ = \ 0$$

in terms of the coefficients $a_n, \ldots, a_0$.

Of course, sometimes you will be able to solve for the solutions of a high-degree polynomial equation (for instance, $x := 1$ is a solution to $X^{100} - 1 = 0$), but this is usually because the polynomial is carefully chosen in order to admit solutions you can find exactly. This is an exceptional case. In general, the only polynomial equations you can expect a guaranteed solution for is degree 1 (linear) and degree 2 (quadratic). If we do encounter higher-degree polynomials in this class, they will be chosen so that it is possible to find exact solutions. However in general we will stick to degree 2 or lower.

Conventions and notation

In this class the natural numbers is the set $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ of nonnegative integers. In particular, we consider 0 to be a natural number.

Unless stated otherwise, the following convention will be in force throughout the entire course:

Global Convention 0.0.3. Throughout, m and n range over $\mathbb{N} = \{0, 1, 2, \ldots\}$.

CHAPTER 1

Linear algebra I

Before commencing with differential equations, we begin with the first of three chapters on linear algebra. This might seem initially unrelated to differential equations (like the one considered in Question 0.0.1) but we will soon find that linear algebra is intimately connected with many of the things we will do with differential equations and it is the best language to explain many different phenomena we will encounter.

1.1. Systems of equations

In this section we will give a crash course in the correct way to completely solve a system of equations (with any number of variables and any number of equations).

Systems of equations. Here is an example of a system of equations:

$$\begin{aligned} 2X + Y &= 1 \\ X - Y &= 1 \end{aligned} \tag{1.1}$$

This is a system of equations with two variables (X and Y) and two equations. A solution to (1.1) is a pair (x, y) of real numbers, such that when we plug in x for X and y for Y, both equations are satisfied. We will recall how one solves (1.1) using what we will call the *naive method*:

Solution to (1.1). First we will multiply the second equation by 2 so that the coefficients on "X" are the same:

$$\begin{aligned} 2X + Y &= 1 \\ 2X - 2Y &= 2 \end{aligned} \tag{1.2}$$

Next we will subtract the first equation from the second equation to eliminate the second "X":

$$\begin{aligned} 2X + Y &= 1 \\ -3Y &= 1 \end{aligned} \tag{1.3}$$

Now we see that $y := -1/3$ is the only value for Y which works. Plugging this into the top equation yields:

$$2X - 1/3 = 1 \quad \text{and thus} \quad X = 2/3.$$

Thus $x := 2/3$ is the only value for X that works. We conclude that $(x, y) = (2/3, -1/3)$ is the *only* solution to (1.1). □

We call this the *naive method* because it relies on observations and *ad hoc* computations. We include it here mainly to jog your memory of how you might have previously learned to solve systems of equations. However, this method quickly becomes burdensome when you consider more variables and more equations. In the

rest of this section, we will introduce the *correct method* you should use to solve these systems. At this point we make the following declaration:

You should never again use the naive method

to solve a system of equations.

Instead you should commit to learning and using the method introduced below. Before we proceed, we will make a few more definitions:

Definition 1.1.1. A **system of equations (with m equations and n variables)** is a system

$$\begin{aligned} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n &= b_1 \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n &= b_2 \\ &\vdots \\ a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n &= b_m \end{aligned} \tag{1.4}$$

where $b_i, a_{ij} \in \mathbb{R}$ for every $i = 1, \ldots, m$ and $j = 1, \ldots, n$. A **solution** to the system (1.4) is an n-tuple $(x_1, x_2, \ldots, x_n)$ of real numbers such that when you plug x_i in for X_i (for each $i = 1, \ldots, n$), each equation is true.

Example 1.1.2. The following system has 3 equations and 4 variables:

$$\begin{aligned} X_1 + 2X_2 - 3X_3 + X_4 &= 6 \\ 2X_1 + X_2 - 2X_3 - X_4 &= 4 \\ 6X_2 + 4X_3 - X_4 &= 4 \end{aligned}$$

and it is easy to check that $(1/3, 4/3, -1, 0)$ is a solution (although there are other solutions as well).

In general the goal will be to find *all* solutions to a system of equations, not just one single solution.

Augmented matrices. Recall that in our solution to the system (1.1) above we first had the system

$$\begin{aligned} 2X + 1Y &= 1 \\ 2X - 2Y &= 2 \end{aligned}$$

which then we transformed into the system

$$\begin{aligned} 2X + 1Y &= 1 \\ 0X - 1Y &= 1. \end{aligned} \tag{1.5}$$

Note also that every symbol colored in red has nothing to do with the specific numbers; the presence and locations of "X", "Y" and "=" is always guaranteed to be exactly the same each time we transform the system. The only thing that matters for each system is what coefficients are in which spot.

This brings us to the first major innovation linear algebra has to offer us for systems of equations: *augmented matrices*. An **augmented matrix** for a system of m

equations in n variables (such as (1.4) above) is a rectangular array with m rows and $n+1$ columns which stores all the coefficients of the system:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

Example 1.1.3. For example, the system

$$\begin{aligned} 3a + 4b + c &= 2 \\ a - 5c &= 3 \end{aligned}$$

has corresponding augmented matrix

$$\left[\begin{array}{ccc|c} 3 & 4 & 1 & 2 \\ 1 & 0 & -5 & 3 \end{array}\right]$$

In other words an augmented matrix is nothing more than a *compact storage device for an entire system of equations.* Whenever you see a system of equations, you should also picture it's augmented matrix, and vice versa.

Henceforth, we will primarily use augmented matrices

for writing systems of equations.

Now we return to the main order of business which is to efficiently solve systems of equations (i.e., determine *all* solutions). Basically, we will learn how to play a game. The game is called **Gaussian Elimination**. The rules of the game are roughly as follows:

(I) There are three legal moves (so-called *elementary row operations*) which we can use to transform one augmented matrix into the next augmented matrix.
(II) When starting out, the first[1] goal is to transform your matrix into *Row Echelon Form.*
(III) After getting to Row Echelon Form, the next goal is to continue to transform your matrix into *Reduced Row Echelon Form.*
(IV) Once the matrix is in Reduced Row Echelon Form, it is very easy to read off all solutions to the original system.

We will study these four things separately in the remainder of this section.

Row operations. Suppose we have an augmented matrix

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

The following **elementary row operations** are the only ways we are allowed to transform this augmented matrix:

(1) (*Row switching*) A row in the matrix can be switched with another row in the matrix. Notation: $R_i \leftrightarrow R_j$

[1]In some linear algebra books and classes, this step is skipped and the goal is to go directly to reduced row echelon form in (III). It's fine if you do it that way, although in general it will take the same amount of work and effort.

(2) (*Row multiplicaton*) A row can be multiplied by a non-zero constant. Notation: $\alpha R_i \to R_i$
(3) (*Row addition*) A row can be replaced with the sum of that row and a multiple of another row. Notation: $R_i + \alpha R_j \to R_i$.

Here is an example of a sequence of three applications of elementary row operations:

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 2 \\ 2 & 4 & 4 & 3 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|c} 2 & 4 & 4 & 3 \\ 0 & 1 & 1 & 2 \end{array}\right] \quad \text{(row switch row 1 and row 2)}$$

$$\xrightarrow{\frac{1}{2}R_1 \to R_1} \left[\begin{array}{ccc|c} 1 & 2 & 2 & 3/2 \\ 0 & 1 & 1 & 2 \end{array}\right] \quad \text{(multiply row 1 by 1/2)}$$

$$\xrightarrow{R_1 - 2R_2 \to R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -5/2 \\ 0 & 1 & 1 & 2 \end{array}\right] \quad \text{(add } -2 \text{ times row 2 to row 1)}$$

Question 1.1.4. *Why are these the only operations allowed?*

Proof. These row operations have the property that they are *reversible*. This means that the set of solutions remains the same in each augmented matrix. Note that if we allowed "multiplication by 0" to be a row operation, then this would have the effect of deleting information in the system and it might introduce additional solutions which are not solutions of the original system (which would be very undesirable). □

Below we will explain how to use these row operations to achieve our objective of solving the original system of equations.

Row echelon form (REF). We will illustrate the entire process with the following example which we will occasionally check back in with:

Example 1.1.5. Find all solutions to the system

$$\begin{aligned} 3X_1 + 6X_2 + 6X_3 &= 24 \\ -6X_1 - 12X_2 - 12X_3 &= -48 \\ 6X_1 + 12X_2 + 10X_3 &= 42 \end{aligned} \tag{1.6}$$

Solution to Example 1.1.5, Part I. The first step is to rewrite the system (1.6) as an augmented matrix:

$$\left[\begin{array}{ccc|c} 3 & 6 & 6 & 24 \\ -6 & -12 & -12 & -48 \\ 6 & 12 & 10 & 42 \end{array}\right] \qquad \square$$

Now we need to know how are we supposed to transform our augmented matrix using the three elementary row operations. First objective is to transform our augmented matrix into *row echelon form*:

Definition 1.1.6. An augmented matrix is in **row echelon form (REF)** if

(1) every row with nonzero entries is above every row with all zeroes (if there are any), and
(2) the leading coefficient of a nonzero row (i.e., the leftmost nonzero entry of that row) is to the right of the leading coefficient of the row above it.

Example 1.1.7. The following augmented matrices are in REF (with the leading coefficients underlined):

$$\left[\begin{array}{cc|c}\underline{4} & 3 & 1\\ 0 & \underline{1} & 2\end{array}\right] \quad \left[\begin{array}{ccc|c}0 & \underline{3} & 1 & 8\end{array}\right] \quad \left[\begin{array}{ccc|c}\underline{1} & 0 & 0 & 1\\ 0 & \underline{1} & 0 & 2\\ 0 & 0 & \underline{1} & 3\\ 0 & 0 & 0 & 0\end{array}\right] \quad \left[\begin{array}{ccc|c}\underline{2} & 3 & 0 & 0\\ 0 & 0 & \underline{1} & 0\end{array}\right] \quad \left[\begin{array}{cc|c}0 & 0 & \underline{2}\\ 0 & 0 & 0\end{array}\right]$$

The following augmented matrices are *not* in REF:

$$\left[\begin{array}{cc|c}0 & 0 & 0\\ 0 & \underline{1} & 1\end{array}\right] \quad \left[\begin{array}{ccc|c}\underline{1} & 0 & 0 & 1\\ 0 & 0 & \underline{1} & 2\\ 0 & \underline{1} & 0 & 3\end{array}\right] \quad \left[\begin{array}{cc|c}0 & \underline{1} & 0\\ 0 & 0 & 0\\ \underline{1} & 0 & 0\end{array}\right]$$

Solution to Example 1.1.5, Part II. Our augmented matrix is not in row echelon form. In particular, the leading coefficients of the second and third row are directly below the leading coefficient of the first row, which is not allowed:

$$\left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ \underline{-6} & -12 & -12 & -48\\ \underline{6} & 12 & 10 & 42\end{array}\right]$$

To fix this, we need to use row addition with the first row to turn the leading -6 and 6 of the second and third row into a zero:

$$\left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ -6 & -12 & -12 & -48\\ 6 & 12 & 10 & 42\end{array}\right] \xrightarrow{R_2+2R_1\to R_2} \left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ 0 & 0 & 0 & 0\\ 6 & 12 & 10 & 42\end{array}\right]$$

$$\xrightarrow{R_3-2R_1\to R_3} \left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ 0 & 0 & 0 & 0\\ 0 & 0 & -2 & -6\end{array}\right]$$

We are still not in row echelon form since we have a row of all zeros above a row with nonzero entries:

$$\left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ \underline{0} & \underline{0} & \underline{0} & \underline{0}\\ 0 & 0 & -2 & -6\end{array}\right]$$

To remedy this, we will switch rows 2 and 3:

$$\xrightarrow{R_2\leftrightarrow R_3} \left[\begin{array}{ccc|c}3 & 6 & 6 & 24\\ 0 & 0 & -2 & -6\\ 0 & 0 & 0 & 0\end{array}\right]$$

We are now in row echelon form and we are done this step. □

Once our augmented matrix is in row echelon form, we can make the following definition:

Definition 1.1.8. Given an augmented in REF, a **pivot** is a leading coefficient in a nonzero row.

For instance, the augmented matrix we arrived at in Example 1.1.5 has two pivots, which we indicate in boxes:

$$\left[\begin{array}{ccc|c}\boxed{3} & 6 & 6 & 24\\ 0 & 0 & \boxed{-2} & -6\\ 0 & 0 & 0 & 0\end{array}\right]$$

Pivots play an important role in Gaussian Elimination. The next step is to take our augmented matrix a little bit further to *reduced row echelon form*.

Reduced row echelon form (RREF). The ultimate goal is to get our augmented matrix into *reduced row echelon form*:

Definition 1.1.9. An augmented matrix is in **reduced row echelon form (RREF)** if

(1) it is in row echelon form (REF),
(2) every pivot is 1, and
(3) every entry above a pivot is 0.

Example 1.1.10. The following augmented matrices are in RREF:

$$\left[\begin{array}{ccc|c} 0 & \boxed{1} & & 0 \end{array}\right] \quad \left[\begin{array}{ccc|c} \boxed{1} & 2 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{array}\right] \quad \left[\begin{array}{cc|c} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 5 \end{array}\right]$$

The following matrices are in REF but *not* RREF:

$$\left[\begin{array}{cc|c} \boxed{4} & 3 & 1 \\ 0 & \boxed{1} & 2 \end{array}\right] \quad \left[\begin{array}{ccc|c} 0 & \boxed{3} & 1 & 8 \end{array}\right] \quad \left[\begin{array}{ccc|c} \boxed{2} & 3 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \end{array}\right]$$

We now continue on with our main example:

Solution to Example 1.1.5, Part III. We see that the augmented matrix we left off with is not in RREF, only REF. This is because the pivots are 3 and -2, not 1 and 1, and also the underlined 6 should be a 0:

$$\left[\begin{array}{ccc|c} \boxed{3} & 6 & \underline{6} & 24 \\ 0 & 0 & \boxed{-2} & -6 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

To remedy this, we use row multiplication to fix the pivot values, and then row addition to get rid of the 6:

$$\xrightarrow{\frac{1}{3}R_1 \to R_1} \left[\begin{array}{ccc|c} 1 & 2 & 2 & 8 \\ 0 & 0 & -2 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{-\frac{1}{2}R_2 \to R_2} \left[\begin{array}{ccc|c} 1 & 2 & 2 & 8 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_1 - 2R_2 \to R_1} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Finally we arrive at RREF. □

Once our augmented matrix is in RREF, it is easy to read off all solutions of the original system.

Getting the final answer from RREF. We will describe how to get the final answer from RREF first in terms of our main example:

Solution to Example 1.1.5, Part IV. First recall that the first three columns correspond to the three variables X_1, X_2, and X_3:

$$\begin{array}{c} \begin{matrix} X_1 & X_2 & X_3 \end{matrix} \\ \left[\begin{array}{ccc|c} \boxed{1} & 2 & 0 & 2 \\ 0 & 0 & \boxed{1} & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Since X_1 and X_3 have pivots in their columns, X_1 and X_3 are called **pivot variables** and the first and third columns are called **pivot columns**. Since X_2 does not have a pivot, it is called a **free variable** and the second column is called a **free column**. Now we read off the solutions using the following steps:

(1) Each free variable is can be any arbitrary value. In this case, we will say that $X_2 = s$, where $s \in \mathbb{R}$ is any number we like.
(2) Next we rewrite the augmented matrix as a system and solve for the pivot variables:

$$\begin{aligned} X_1 + 2X_2 &= 2 \\ X_3 &= 3 \\ 0 &= 0 \end{aligned}$$

which simplifies to:

$$\begin{aligned} X_1 &= 2 - 2s \\ X_3 &= 3. \end{aligned}$$

We now have our final answer: every solution is of the form:

$$\begin{aligned} X_1 &= 2 - 2s \\ X_2 &= s \\ X_3 &= 3, \end{aligned}$$

where $s \in \mathbb{R}$ can be any number. We write the set of all solutions as follows:

$$\big\{(2 - 2s, s, 3) : s \in \mathbb{R}\big\}$$

This way of describing the set of solutions is often called **parametric form** because it describes the solutions in terms of the free parameter s. Notice that there are infinitely many solutions, since there are infinitely many values of s. To get specific solutions, you can just choose values of s. For instance, $s := 0$ yields the solution $(2, 0, 3)$, whereas $s := 10$ yields the solution $(-18, 10, 3)$. □

Example 1.1.11. In this example we will see what to do with 2 free variables. Suppose we are given some system which has the following RREF:

$$\begin{array}{c} \begin{matrix} X_1 & X_2 & X_3 & X_4 \end{matrix} \\ \left[\begin{array}{cccc|c} 0 & \boxed{1} & 2 & 0 & 7 \\ 0 & 0 & 0 & \boxed{1} & 8 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Then we have two free variables X_1 and X_3, so we need to introduce two parameters $s, t \in \mathbb{R}$ and set $X_1 = s$ and $X_3 = t$. Then the system becomes:

$$\begin{aligned} X_2 + 2X_3 &= 7 \\ X_4 &= 8 \end{aligned}$$

and so the general solution is:

$$\begin{aligned} X_1 &= s \\ X_2 &= -2t + 7 \\ X_3 &= t \\ X_4 &= 8 \end{aligned}$$

where $s, t \in \mathbb{R}$ are arbitrary. We can write the set of solutions in parametric form as follows:

$$\big\{(s, -2t + 7, t, 8) : s, t \in \mathbb{R}\big\}$$

Note that to get a specific solution, we are free to choose any s and any t we like. For instance, $s = 1, t = 0$ gives the solution $(1, 7, 0, 8)$ whereas $s = 0, t = 1$ gives the solution $(0, 5, 1, 8)$.

Example 1.1.12. We will give an example of a system with no solutions. Suppose we are given a system with the following RREF:

$$\begin{array}{c} \begin{matrix} X_1 & X_2 & \end{matrix} \\ \left[\begin{array}{cc|c} \boxed{1} & 2 & 0 \\ 0 & 0 & \boxed{1} \end{array}\right] \end{array}$$

Converting this augmented matrix back to a system of equations yields:

$$\begin{aligned} 1X_1 + 2X_2 &= 0 \\ 0X_1 + 0X_2 &= 1 \end{aligned}$$

We claim there cannot be any solutions. Indeed, if say (x_1, x_2) is a solution, then this would mean it satisfies both equations, in particular, the bottom equation. Then $0x_1 + 0x_2 = 1$, i.e., $0 = 1$. However this is always false.

We conclude this section with some more terminology and some general facts:

Definition 1.1.13. We say that a system of equations is **consistent** if it has at least one solution, and we say a system of equations is **inconsistent** if it does not have any solutions.

Fact 1.1.14. Given a system of equations, exactly one of the following three things will happen:

(1) The system has zero solutions (i.e., it is inconsistent). This happens when the RREF contains a row of the form

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right]$$

because this corresponds to the equation $0 = 1$ which can never be true.
(2) The system has exactly one solution. This happens when the system is consistent and there are no free variables in the RREF.
(3) The system has infinitely many solutions. This happens when the system is consistent and there is at least one free variable in the RREF.

In fact, all 3 of the above cases can be determined once you're in REF. If you only care about *how many* solutions there are (and not what exactly they are), then you can just stop once you get to REF. This is one of the benefits of going through the REF on your way to RREF.

Here are some cardinal rules to always follow:

(1) Always recopy the entire augmented matrix in each step, even if you are copying a row of zeros. It is important that the size of the augmented matrix (3×4 in our example) does not change.
(2) Always denote which row operation you are performing in each step.
(3) Always do one row operation at a time, at least when you are starting out. If you attempt to do multiple row operations in one step then this can lead to errors.

Remark 1.1.15. Given a system of equations, we take it to RREF and obtain the set of solutions for the *original system we started out with*. However, this is actually the set of solutions for *every system we encountered along the way*. This is because the RREF of the original system also works as the RREF for every intermediate system.

Geometric interpretation. When you are solving systems of equations, it is good to keep in mind the underlying geometric interpretation. Recall that a linear equation in two variables:

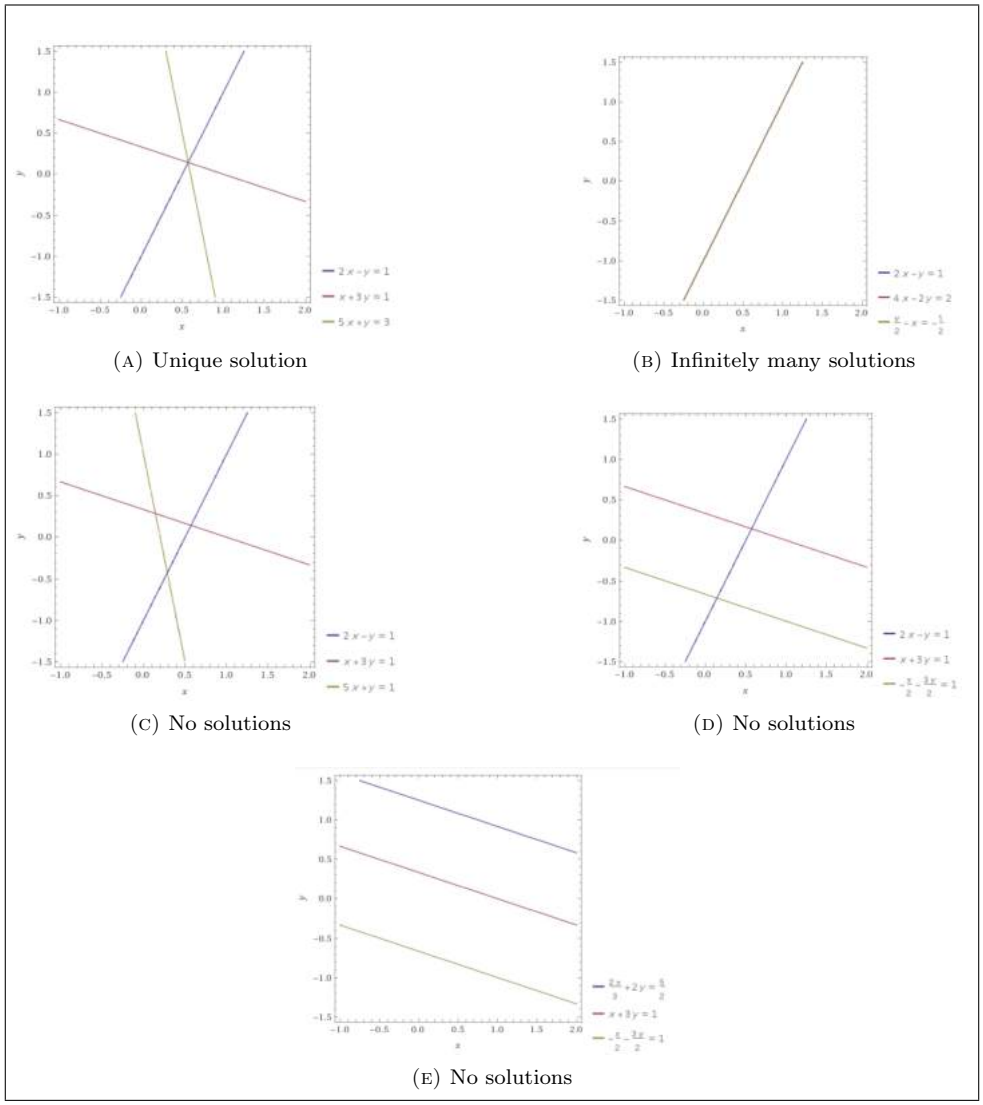
$$2x + 3y = 1$$

can also be viewed as an equation for a line in the plane ($y = -\frac{2}{3}x + \frac{1}{3}$). Thus, a system of linear equations:

$$\begin{aligned} 2x + 3y &= 1 \\ 5x + 7y &= 2 \\ x + y &= 3 \end{aligned}$$

is really asking us to find all points (x, y) in the plane which are part of all three lines, i.e., we want to know where do these three lines intersect, if at all. If we are consider three variables, then we are asking where do multiple planes simultaneously intersect, if at all. For more than 3 variables, we are asking where do higher-dimensional hyper-planes intersect in higher-dimensional euclidean space (something difficult to visualize).

In Figure 1.1, we consider five systems of equations, where each one has two variables and three equations. You can see that there are different ways that the cases *no solutions*, *exactly one solution*, and *infinitely many solutions* can arise.



(A) Unique solution

(B) Infinitely many solutions

(C) No solutions

(D) No solutions

(E) No solutions

FIGURE 1.1. Possible intersections of three lines in a plane

Some specifics about terminology. In this section, we have only been working with *augmented matrices*, for instance

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right] \tag{1.7}$$

An augmented matrix is just a special example of a *matrix* with a vertical bar which superficially separates the columns. A **matrix** (with m rows and n columns) is a rectangular array of numbers:

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

For instance, the augmented matrix (1.7) is considered a 2×3 matrix. When discussing an augmented matrix, we will always consider every column as part of the augmented matrix. If we want to refer only to the entries to the left of the vertical bar:

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$$

this will be referred to as the **coefficient matrix** (of the linear system).

Definition 1.1.16. Here are some precise definitions summarized:

(1) Given a matrix, a **leading entry** of a row is the leftmost nonzero entry (if there is one). In the following matrices, we underline the leading entries:

$$\begin{bmatrix} \underline{2} & 3 & 0 \\ 0 & \underline{2} & 1 \\ \underline{1} & 0 & 2 \end{bmatrix} \quad \left[\begin{array}{cc|c} \underline{1} & 2 & 0 \\ 0 & 0 & \underline{1} \end{array}\right]$$

(2) If a matrix is in REF, then the leading entries are also called **pivots**. The following matrices are in REF and the pivots are in boxes:

$$\left[\begin{array}{ccc|c} \boxed{2} & 3 & 5 & 4 \\ 0 & 0 & \boxed{7} & 1 \\ 0 & 0 & 0 & \boxed{2} \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{bmatrix} 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(3) If a matrix is *not* in REF, then we choose not to define what a pivot is. In this class we will only discuss "pivots" in the context of Gaussian Elimination and only allow ourselves to refer to "the pivots of a matrix" if we know the matrix is already in REF. For all matrices, the expression "leading entry" will always make sense, regardless of whether the matrix is in REF or not.

(4) We define the **rank** of a matrix to be the number of pivots any REF of that matrix has (it will be the same number even though there could be many different REFs).

Question 1.1.17. *Why are we reluctant to call leading coefficients in a non-REF matrix "pivots"?*

Answer 1.1.18. In general, a *pivot* (noun) is something that you *pivot* (verb) around. Given a nonzero entry of a matrix, to **pivot** around that entry means to use elementary row operations to turn that entry into a 1 and then use it to turn the other entries in that column into 0. In the following example, we pivot around the boxed entry (for no particular reason other than to show an example of "pivoting"):

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & \boxed{2} & 2 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{\frac{1}{2}R_2 \to R_2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \boxed{1} & 1 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{R_1 - R_2 \to R_1}$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & \boxed{1} & 1 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{R_3 - 3R_1 \to R_3} \begin{bmatrix} 0 & 0 & 0 \\ 1 & \boxed{1} & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Since this is what "pivoting" means, we define *pivots* so that in Gaussian Elimination we are essentially *pivoting around the pivots*. We do not pivot around the leading entries which are not pivots. Furthermore, there are other algorithms in linear algebra besides Gaussian Elimination (for instance, the *Simplex Algorithm*[2]) where you pivot around entries which are not leading coefficients. Thus, you shouldn't get too attached to the idea "pivot means leading entry".

Given the above discussion, we can now recast some of the above facts in more detail:

Fact 1.1.19. Suppose we are considering a system of equations which has augmented matrix:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

and coefficient matrix:

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

(1) The following are equivalent:
   (a) the system has no solutions,
   (b) the system is inconsistent,
   (c) an REF of the augmented matrix has a row of the form
   $$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & \neq 0 \end{array}\right],$$
   (d) the RREF of the augmented matrix has a row of the form
   $$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right],$$
   (e) an REF of the augmented matrix has a pivot in the last column,
   (f) the RREF of the augmented matrix has a pivot in the last column,
   (g) the rank of the coefficient matrix is not equal to the rank of the entire augmented matrix.

(2) Suppose the system is consistent. Then the following are equivalent:
   (a) the system has exactly one solution,
   (b) every variable is a pivot variable,
   (c) there are no free variables,
   (d) the rank of the augmented matrix is equal to the number of columns in the coefficient matrix (= number of variables).

(3) Suppose the system is consistent. Then the following are equivalent:
   (a) the system has infinitely many solutions,
   (b) at least one variable is a free variable,
   (c) the rank of the augmented matrix is less than the number of columns in the coefficient matrix (i.e., less than the number of variables).

[2] `https://en.wikipedia.org/wiki/Simplex_algorithm`

1.2. Application: partial fractions

In this section, we revisit the powerful method of *partial fractions*, viewed as an application of linear systems.

Case I: distinct linear factors. Suppose we want to integrate the rational function:

$$\frac{3x+4}{x^3-3x^2+2x}$$

To do this, we must first factor the denominator polynomial: $x^3-3x^2+2x = (x-0)(x-1)(x-2)$. Since there are no (strictly) complex roots, this polynomial factors into linear factors (with real roots). Also, for this polynomial, every linear factor is distinct (occurs with multiplicity one). Thus, the general form of the partial fraction decomposition is:

$$\frac{3x+4}{x(x-1)(x-2)} \;=\; \frac{A}{x}+\frac{B}{x-1}+\frac{C}{x-2},$$

where $A, B, C \in \mathbb{R}$ are three unknown real numbers we need to solve for. Clearing denominators yields:

$$3x+4 \;=\; A(x-1)(x-2)+Bx(x-2)+C(x-1)(x-2)$$

This equality is to be interpreted as: for every possible real number $x \in \mathbb{R}$, when you plug x into both the lefthand side and the righthand side, you should get a true equality of two numbers. We will use this observation and plug in three carefully chosen numbers to see what they give us:

- $(x=0)$ In this case, the equation becomes $4 = 2A$
- $(x=1)$ In this case, the equation becomes $7 = -B$
- $(x=2)$ In this case, the equation becomes $10 = 2C$

Thus, we have arrived at a (easy) system of equations:

$$\begin{aligned} 2A &= 4 \\ -B &= 7 \\ 2C &= 10. \end{aligned}$$

We can solve this system using Gaussian Elimination:

$$\left[\begin{array}{ccc|c} 2 & 0 & 0 & 4 \\ 0 & -1 & 0 & 7 \\ 0 & 0 & 2 & 10 \end{array}\right] \xrightarrow{\frac{1}{2}R_1\to R_1,\,-R_2\to R_2,\,\frac{1}{2}R_3\to R_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

This gives us the unique solution $(A, B, C) = (2, -7, 5)$. We conclude that

$$\frac{3x+4}{x^3-3x^2+2x} \;=\; \frac{2}{x}-\frac{7}{x-1}+\frac{5}{x-2}$$

is our desired partial fraction decomposition. The rational function can now be integrated using the logarithm.

Case II: repeated linear factors. Suppose now we wish to decompose

$$\frac{5x^3+6x^2+7x+8}{x^4-2x^3+x^2}$$

We are able to factor the denominator as $x^4-2x^3+x^2=x^2(x-1)^2$. We see that there are two linear factors, each one with multiplicity two. Thus the general form of the partial fraction decomposition is

$$\frac{5x^3+6x^2+7x+8}{x^2(x-1)^2} = \frac{A}{x}+\frac{B}{x^2}+\frac{C}{x-1}+\frac{D}{(x-1)^2}$$

where $A,B,C,D\in\mathbb{R}$ are four unknown real numbers we need to solve for (the rule is, for each multiplicity of a linear factor, you get another term in the expansion and another variable). First we cross-multiply so that we have an equality of polynomials, then we rewrite the righthand side as a single polynomial:

$$\begin{aligned} 5x^3+6x^2+7x+8 &= Ax(x-1)^2+B(x-1)^2+Cx^2(x-1)+Dx^2 \\ &= A(x^3-2x^2+x)+B(x^2-2x+1)+C(x^3-x^2)+Dx^2 \\ &= (A+C)x^3+(-2A+B-C+D)x^2+(A-2B)x+B. \end{aligned}$$

Next, we use the important observation that two polynomials are the same if and only if they have the same degree and the corresponding coefficients are the same. Thus the above equality of polynomials yields the system:

$$\begin{aligned} A+C &= 5 \\ -2A+B-C+D &= 6 \\ A-2B &= 7 \\ B &= 8. \end{aligned}$$

We can now solve the system using Gaussian Elimination:

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 5 \\ -2 & 1 & -1 & 1 & 6 \\ 1 & -2 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 & 8 \end{array}\right] \xrightarrow{\text{to RREF (steps omitted)}} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 23 \\ 0 & 1 & 0 & 0 & 8 \\ 0 & 0 & 1 & 0 & -18 \\ 0 & 0 & 0 & 1 & 26 \end{array}\right]$$

We find that the unique solution is $(A,B,C,D)=(23,8,-18,26)$. Thus the desired partial fraction decomposition is

$$\frac{5x^3+6x^2+7x+8}{x^4-2x^3+x^2} = \frac{23}{x}+\frac{8}{x^2}-\frac{18}{x-1}+\frac{26}{(x-1)^2}$$

Case III: irreducible quadratic factors. Technically speaking, if you are comfortable working with complex numbers and complex-valued functions, then you only ever have to consider factorizations of the denominator into linear factors. However, for various reasons it is convenient to have a method of partial fraction decomposition which does not require us to ever leave the realm of real numbers. For instance, for the following rational function

$$\frac{10x^2+11x+12}{(x^2+1)(x+1)}$$

we *could* factor the denominator into linear factors

$$(x^2+1)(x+1) = (x+i)(x-i)(x+1),$$

and then proceed as in Case I (which we'll do below just to prove a point). However, we can just as easily keep the quadratic factor x^2+1 as is in our computation. Since in general the number of unknowns in a partial fraction decomposition must be equal to the degree of the denominator polynomial, the quadratic factor has to contribute two unknowns to the general form:

$$\frac{10x^2+11x+12}{(x^2+1)(x+1)} = \frac{Ax+B}{x^2+1}+\frac{C}{x+1}$$

We now proceed as in Case II by clearing denominators and getting an equality of two polynomials:

$$\begin{aligned} 10x^2+11x+12 &= (Ax+B)(x+1)+C(x^2+1) \\ &= (A+C)x^2+(A+B)x+(B+C) \end{aligned}$$

This gives us a system of equations:

$$\begin{aligned} A+C &= 10 \\ A+B &= 11 \\ B+C &= 12 \end{aligned}$$

which we can solve using Gaussian Elimination

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 10 \\ 1 & 1 & 0 & 11 \\ 0 & 1 & 1 & 12 \end{array}\right] \xrightarrow{\text{to RREF (steps omitted)}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 9/2 \\ 0 & 1 & 0 & 13/2 \\ 0 & 0 & 1 & 11/2 \end{array}\right]$$

This gives us the desired partial fraction expansion:

$$\frac{10x^2+11x+12}{(x^2+1)(x+1)} = \frac{9x+13}{2(x^2+1)}+\frac{11}{2(x+1)}$$

We can check our work by re-doing the decomposition with complex numbers:

$$\frac{10x^2+11x+12}{(x+i)(x-i)(x+1)} = \frac{A}{x+i}+\frac{B}{x-i}+\frac{C}{x+1}$$

Cross-multiplying yields

$$10x^2+11x+12 = A(x-i)(x+1)+B(x+i)(x+1)+C(x-i)(x+i)$$

Now we plug in the three denominator roots to get linear equations for the unknowns:

- $(x=-i)$ In this case, the equation becomes $2-11i=(-2-2i)A$
- $(x=i)$ In this case, the equation becomes $2+11i=(-2+2i)B$
- $(x=-1)$ In this case, the equation becomes $11=2C$

This yields the system:

$$\begin{aligned} (-2-2i)A &= 2-11i \\ (-2+2i)B &= 2+11i \\ 2C &= 11 \end{aligned}$$

which we can solve with Gaussian Elimination:

$$\left[\begin{array}{ccc|c} -2-2i & 0 & 0 & 2-11i \\ 0 & -2+2i & 0 & 2+11i \\ 0 & 0 & 2 & 11 \end{array}\right] \xrightarrow{\text{to RREF (steps omitted)}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & (9+13i)/4 \\ 0 & 1 & 0 & (9-13i)/4 \\ 0 & 0 & 1 & 11/2 \end{array}\right]$$

This yields the desired partial fraction decomposition:

$$\frac{10x^2+11x+12}{(x+i)(x-i)(x+1)} = \frac{9+13i}{4(x+i)} + \frac{9-13i}{4(x-i)} + \frac{11}{2(x+1)}$$

Finally, to pull this decomposition back into the realm of real numbers, we add the first two fractions together (since those two correspond to a conjugate pair of roots):

$$\begin{aligned}\frac{13-9i}{4(x+i)} + \frac{9-13i}{4(x-i)} + \frac{11}{2(x+1)} &= \frac{(9+13i)(x-i)+(9-13i)(x+i)}{4(x+i)(x-i)} + \frac{11}{2(x+1)} \\ &= \frac{9x+13}{2(x^2+1)} + \frac{11}{2(x+1)}\end{aligned}$$

This shows that working with complex numbers gives the same decomposition.

Case IV: repeated quadratic factors. Finally, we arrive at perhaps the most involved case: *repeated quadratic factors*. However, the method here is really just the same as the methods in Cases II and III provided you know the rule for the general form. Here is an example:

$$\frac{6x^3+7x^2+8x+9}{(x^2+x+1)^2}$$

Since the quadratic factor x^2+x+1 has multiplicity two, it has two show up twice in the decomposition. Since the total number of unknowns needs to be four (= degree of denominator polynomial), each occurrence of the quadratic factor has to have two unknowns:

$$\frac{6x^3+7x^2+8x+9}{(x^2+x+1)^2} = \frac{Ax+B}{x^2+x+1} + \frac{Cx+D}{(x^2+x+1)^2}$$

Just as before, we cross-multiply and get an equality of polynomials:

$$\begin{aligned}6x^3+7x^2+8x+9 &= (Ax+B)(x^2+x+1)+Cx+D \\ &= Ax^3+(A+B)x^2+(A+B+C)x+(B+D)\end{aligned}$$

Equating the two polynomials gives us the system of equations:

$$\begin{aligned}A &= 6 \\ A+B &= 7 \\ A+B+C &= 8 \\ B+D &= 9\end{aligned}$$

which we can solve using Gaussian Elimination:

$$\left[\begin{array}{cccc|c}1&0&0&0&6\\1&1&0&0&7\\1&1&1&0&8\\0&1&0&1&9\end{array}\right] \xrightarrow{\text{to RREF (steps omitted)}} \left[\begin{array}{cccc|c}1&0&0&0&6\\0&1&0&0&1\\0&0&1&0&1\\0&0&0&1&8\end{array}\right]$$

This gives us the desired partial fraction decomposition:

$$\frac{6x^3+7x^2+8x+9}{(x^2+x+1)^2} = \frac{6x+1}{x^2+x+1} + \frac{x+8}{(x^2+x+1)^2}$$

CHAPTER 2

Calculus review

In this section we will summarize all the important definitions and results from calculus. In general we will state these results for arbitrary nice functions, for summary of calculus results pertaining to special elementary functions, see Appendix A. First, some terminology which will simplify some things. Given the set of real numbers $\mathbb{R}$, we artificially adjoin two new symbols $+\infty$ and $-\infty$ to serve as convenient bookends of the ordering. More specifically:

Definition 2.0.1. Define the **extended real numbers** to be the set $\mathbb{R}_{\pm\infty} := \mathbb{R} \cup \{-\infty, +\infty\}$. We extend the ordering on $\mathbb{R}$ to all of $\mathbb{R}_{\pm\infty}$ by declaring:

$$-\infty \leq a \leq +\infty \quad \text{for every } a \in \mathbb{R}_{\pm\infty}.$$

Unless we state otherwise, we do not extend the arithmetic operations $+, \cdot$ on $\mathbb{R}$ to include $\pm\infty$. It is important to realize the new elements $\pm\infty$ are *not* numbers and there is not supposed to be anything super deep or special about adjoining $\pm\infty$ to our real line. We primarily introduce it because it makes certain commonly occurring statements and expressions shorter.

For instance, we can define *bounded intervals* and *unbounded intervals* with uniform notation. Given $a, b \in \mathbb{R}$ such that $a < b$, an **interval** is a set of one of the following forms:

$$\begin{aligned}
(a,b) &:= \{x \in \mathbb{R} : a < x < b\} \\
[a,b) &:= \{x \in \mathbb{R} : a \leq x < b\} \\
(a,b] &:= \{x \in \mathbb{R} : a < x \leq b\} \\
[a,b] &:= \{x \in \mathbb{R} : a \leq x \leq b\} \\
(a,+\infty) &:= \{x \in \mathbb{R} : a < x\} \\
[a,+\infty) &:= \{x \in \mathbb{R} : a \leq x\} \\
(-\infty,b) &:= \{x \in \mathbb{R} : x < b\} \\
(-\infty,b] &:= \{x \in \mathbb{R} : x \leq b\} \\
(-\infty,+\infty) &:= \mathbb{R}
\end{aligned}$$

Intervals of the form (a,b), $[a,b)$, $(a,b]$, $[a,b]$ are called **bounded intervals**. Intervals of the form $(a,+\infty)$, $[a,+\infty)$, $(-\infty,b)$, $(-\infty,b]$, $(-\infty,+\infty)$ are called **unbounded intervals**. Intervals of the form (a,b), $(a,+\infty)$, $(-\infty,b)$, $(-\infty,+\infty)$ are call **open intervals**. Intervals of the form $[a,b]$, $[a,+\infty)$, $(-\infty,b]$, $(-\infty,+\infty)$ are called **closed intervals**.

Of course, intervals are not the only types of subsets of $\mathbb{R}$ which naturally arise in this class. For instance, the natural domain of the tangent function is not an

interval, but instead a union of intervals:

$$\begin{aligned}\text{domain}(\tan t) &= \{t \in \mathbb{R} : t \neq \pi/2 + \pi k \text{ for every } k \in \mathbb{Z}\} \\ &= \bigcup_{k\in\mathbb{Z}} \Big(\frac{\pi}{2} + \pi k, \frac{\pi}{2} + \pi(k+1)\Big)\end{aligned}$$

In order to avoid too many technicalities, we will consider a subset $D \subseteq \mathbb{R}$ to be *nice* if can show up as the true domain of some function one would encounter in freshman calculus. To be specific:

Definition 2.0.2. We call a set $D \subseteq \mathbb{R}$ **nice** if it is an interval or a union of a sequence of intervals, i.e., if there exists a sequence of intervals $I_0, I_1, I_2, \ldots$ such that

$$D = \bigcup_{n\geq 0} I_n$$

In general we will always restrict our attention to functions with nice domains, with the domain of the tangent function being representative of the worst type of nice domain. If you find the definition of *nice* too technical, then surprisingly very little is lost if you just interpret the adjective *nice* in the colloquial sense. Really, these things won't matter too much for this class (since you're being graded primarily on learning how to do calculations), but we introduce this terminology anyway so that way in these notes we can still restrict ourselves to making statements which are literally true in a mathematical sense, without being overly abstract and technical.

In the exposition we will occasionally refer to *elementary functions*. We don't mean anything too precise by this, although you can take the following as a rough definition:

Rough Definition 2.0.3. An **elementary function** $f : D \to \mathbb{R}$ is any function constructed from the following operations:

(1) arithmetic operations: $+, -, \cdot, /$
(2) algebraic operations such as taking nth roots
(3) composition of functions
(4) the exponential $\exp : \mathbb{R} \to \mathbb{R}$ and logarithm $\ln : [0, +\infty) \to \mathbb{R}$,
(5) the trigonometric functions $\sin, \cos, \tan$
(6) the inverse trigonometric functions $\arcsin, \arccos, \arctan$

In other words, an *elementary function* is the type of function which shows up in freshman calculus.

2.1. Limits

In this section D is a nice set. We will review the definition and rules for computing limits. Recall that sometimes, even if a function $f : (a, b) \to \mathbb{R}$ is defined on an open interval (a, b), it sometimes still makes sense to ask what is the limit of $f(x)$ as $x \to a$, i.e., $\lim_{x\to a} f(x)$, even though f is not defined at a. This makes sense because a is an endpoint of (a, b), so there are points in (a, b) which are arbitrarily closed to a. In general we will consider functions $f : D \to \mathbb{R}$ where the domain D is a nice set. Before we define *limit*, it first makes sense to define what is the set of all points which it might make sense to take the limit to.

Definition 2.1.1. Define the **closure of** D to be the slightly larger set $\mathrm{cl}(D) \supseteq D$ defined such that for every $\alpha \in \mathbb{R}_{\pm\infty}$, we say that $\alpha \in \mathrm{cl}(D)$ if there exists $x \in \mathbb{R}$ such that either:

(1) $x < \alpha$ and $(x, \alpha) \subseteq D$, or
(2) $\alpha < x$ and $(\alpha, x) \subseteq D$.

In particular, if $\alpha \in D$, then $\alpha \in \mathrm{cl}(D)$. In other words, $\mathrm{cl}(D)$ is the same thing as D plus all the endpoints of the intervals which define D. For example:

$$\begin{aligned} \mathrm{cl}\big((1,2]\big) &= [1,2] \\ \mathrm{cl}\big((-1,0)\cup(0,1]\big) &= [-1,1] \\ \mathrm{cl}\big(\mathrm{domain}(\tan t)\big) &= \mathbb{R} \end{aligned}$$

We can now define in one definition every type of limit of a function encountered in freshman calculus:

Definition 2.1.2. Suppose $f : D \to \mathbb{R}$ is a function with nice domain D. Suppose $\alpha \in \mathrm{cl}(D)$ and $L \in \mathbb{R}_{\pm\infty}$. We say the limit of f as x approaches α exists and is equal to L, notation:

$$\lim_{x\to\alpha} f(x) = L$$

if one of the following is satisfied (depending on whether $\alpha, L = \pm\infty$ or not):

(1) $(\alpha, L \in \mathbb{R})$ for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in D$, if $0 < |x - \alpha| < \delta$, then $\big|f(x) - L\big| < \epsilon$.
(2) $(\alpha = +\infty, L \in \mathbb{R})$ for every $\epsilon > 0$, there exists $M \in \mathbb{R}$ such that for all $x \in D$, if $M < x$, then $\big|f(x) - L\big| < \epsilon$.
(3) $(\alpha = -\infty, L \in \mathbb{R})$ for every $\epsilon > 0$, there exists $M \in \mathbb{R}$ such that for all $x \in D$, if $x < M$, then $\big|f(x) - L\big| < \epsilon$.
(4) $(\alpha \in \mathbb{R}, L = +\infty)$ for every $M \in \mathbb{R}$, there exists $\delta > 0$ such that for all $x \in D$, if $0 < |x - \alpha| < \delta$, then $M < f(x)$.
(5) $(\alpha = L = +\infty)$ for every $M \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that for all $x \in D$, if $N < x$, then $M < f(x)$.
(6) $(\alpha = -\infty, L = +\infty)$ for every $M \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that for all $x \in D$, if $x < N$, then $M < f(x)$.
(7) $(\alpha \in \mathbb{R}, L = -\infty)$ for every $M \in \mathbb{R}$, there exists $\delta > 0$ such that for all $x \in D$, if $0 < |x - \alpha| < \delta$, then $f(x) < M$.
(8) $(\alpha = +\infty, L = -\infty)$ for every $M \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that for all $x \in D$, if $N < x$, then $f(x) < M$.
(9) $(\alpha = L = -\infty)$ for every $M \in \mathbb{R}$, there exists $N \in \mathbb{R}$ such that for all $x \in D$, if $x < N$, then $f(x) < M$.

In general, for this class if and when we compute limits, we will not use directly Definition 2.1.2. Instead we will use known formulas for limits of special functions (see Appendix A) along with various limit laws, including facts about continuity.

Here is the general limit law for sums of limits:

Addition Limit Law 2.1.3. *Suppose $f, g : D \to \mathbb{R}$ are functions where D is a nice domain. Further suppose $\alpha \in \mathrm{cl}(D)$ and the limits*

$$\lim_{x\to\alpha} f(x) = L_f \quad \textit{and} \quad \lim_{x\to\alpha} g(x) = L_g$$

exist with $L_f, L_g \in \mathbb{R}_{\pm\infty}$. Then:

(1) *if* $L_f, L_g \in \mathbb{R}$, *then*
$$\lim_{x\to\alpha}(f+g)(x) \;=\; L_f + L_g$$
(2) *if* $L_f = +\infty$ *and* $L_g \neq -\infty$, *or* $L_g = +\infty$ *and* $L_f \neq -\infty$, *then*
$$\lim_{x\to\alpha}(f+g)(x) \;=\; +\infty$$
(3) *if* $L_f = -\infty$ *and* $L_g \neq +\infty$, *or* $L_g = -\infty$ *and* $L_f \neq +\infty$, *then*
$$\lim_{x\to\alpha}(f+g)(x) \;=\; -\infty$$
(4) *if* $L_f = +\infty$ *and* $L_g = -\infty$, *or* $L_f = -\infty$ *and* $L_g = +\infty$, *then more subtle investigation is needed (l'Hôpital's rule).*

Here is the general limit law for products of limits:

Product Limit Law 2.1.4. *Suppose* $f, g : D \to \mathbb{R}$ *are functions where* D *is a nice domain. Further suppose* $\alpha \in \mathrm{cl}(D)$ *and the limits*
$$\lim_{x\to\alpha} f(x) \;=\; L_f \quad \textit{and} \quad \lim_{x\to\alpha} g(x) \;=\; L_g$$
exist with $L_f, L_g \in \mathbb{R}_{\pm\infty}$. *Then:*

(1) *if* $L_f, L_g \in \mathbb{R}$, *then*
$$\lim_{x\to\alpha}(f\cdot g)(x) \;=\; L_f \cdot L_g$$
(2) *if one of the following is true:*
   (a) $L_f = +\infty$ *and* $L_g > 0$
   (b) $L_f = -\infty$ *and* $L_g < 0$
   (c) $L_f < 0$ *and* $L_g = -\infty$
   (d) $L_f > 0$ *and* $L_g = +\infty$

   then
$$\lim_{x\to\alpha}(f\cdot g)(x) \;=\; +\infty$$
(3) *if one of the following is true:*
   (a) $L_f = -\infty$ *and* $L_g > 0$
   (b) $L_f = +\infty$ *and* $L_g < 0$
   (c) $L_f < 0$ *and* $L_g = +\infty$
   (d) $L_f > 0$ *and* $L_g = -\infty$

   then
$$\lim_{x\to\alpha}(f\cdot g)(x) \;=\; -\infty$$
(4) *if one of the following is true:*
   (a) $L_f = 0$ *and* $L_g = \pm\infty$
   (b) $L_f = \pm\infty$ *and* $L_g = 0$,

   then more subtle investigation is needed (l'Hôpital's rule).

Finally, here is the general limit law for quotients of functions:

Quotient Limit Law 2.1.5. *Suppose* $f, g : D \to \mathbb{R}$ *are functions where* D *is a nice domain. Define the set:*
$$D' \;:=\; \big\{x \in D : g(x) \neq 0\big\} \;\subseteq\; \mathbb{R}.$$
Assume that D' *is also nice (for us it always will be) and suppose for* $\alpha \in \mathrm{cl}(D') \subseteq \mathrm{cl}(D)$ *the limits*
$$\lim_{x\to\alpha} f(x) \;=\; L_f \quad \textit{and} \quad \lim_{x\to\alpha} g(x) \;=\; L_g$$

exist with $L_f, L_g \in \mathbb{R}_{\pm\infty}$. Then for the quotient function:

$$\frac{f}{g} : D' \to \mathbb{R}$$

we have:

(1) if $L_f \in \mathbb{R}$, and $L_g \in \mathbb{R}$ and $L_g \neq 0$, we have

$$\lim_{x\in\alpha} \left(\frac{f}{g}\right)(x) \;=\; \frac{L_f}{L_g}$$

(2) if $L_f \neq \pm\infty$ and $L_g = \pm\infty$, we have

$$\lim_{x\in\alpha} \left(\frac{f}{g}\right)(x) \;=\; 0$$

(3) if $L_f = +\infty$ and $L_g > 0$, or $L_f = -\infty$ and $L_g < 0$, then

$$\lim_{x\in\alpha} \left(\frac{f}{g}\right)(x) \;=\; +\infty$$

(4) if $L_f = +\infty$ and $L_g < 0$, or $L_f = -\infty$ and $L_g > 0$, then

$$\lim_{x\in\alpha} \left(\frac{f}{g}\right)(x) \;=\; +\infty$$

(5) otherwise a more subtle investigation is needed (l'Hôpital's rule).

Multivariable functions. We will also need to occasionally consider functions with multiple variables:

$$F(t, y) \quad \text{where} \quad F : D \to \mathbb{R} \quad D \subseteq \mathbb{R}^2 \text{ is a subset of the } ty\text{-plane}$$

We will not attempt to define what a "nice" subset D of the plane is, although most of our domains will be of the form $D = I \times J$, where I and J are intervals (such a set could be a called a **rectangle**). Ultimately, we will not be in the business of computing limits of multivariable functions in this class, although here is a definition anyway:

Definition 2.1.6. Suppose $F : D \to \mathbb{R}$ is a two-variable function with domain $D \subseteq \mathbb{R}^2$ a nice subset of the ty-plane (think $D = I \times J$, a rectangle). Given a real number $L \in \mathbb{R}$ and a point $(t_0, y_0) \in D$, we say that L is the **limit** of F as (t, y) approaches (t_0, y_0), notation:

$$\lim_{(t,y)\to(t_0,y_0)} F(t, y) \;=\; L$$

if: for every $\epsilon > 0$, there exists $\delta > 0$, such that for every $(t, y) \in D$,

$$\text{if } 0 < \sqrt{(t - t_0)^2 + (y - y_0)^2} < \delta, \text{ then } \left|F(t, y) - L\right| < \epsilon.$$

Even if we were computing multivariable limits in this class, we would rarely use Definition 2.1.6 directly and instead rely on limit laws and facts about continuity.

Limit Laws for Multivariable Functions 2.1.7. *Suppose $F, G : D \to \mathbb{R}$ are two two-variable functions defined on a nice domain and suppose $(t_0, y_0) \in D$. Furthermore, suppose $L_F, L_G \in \mathbb{R}$ are such that*

$$\lim_{(t,y)\to(t_0,y_0)} F(t, y) \;=\; L_F \quad \textit{and} \quad \lim_{(t,y)\to(t_0,y_0)} G(t, y) \;=\; L_G.$$

Then:

(1) $\lim_{(t,y)\to(t_0,y_0)}(F+G)(t,y) = L_F + L_G$,
(2) $\lim_{(t,y)\to(t_0,y_0)}(F\cdot G)(t,y) = L_F \cdot L_G$.

Furthermore, define

$$D' \;:=\; \big\{(t,y)\in D : G(t,y)\neq 0\big\}$$

then, if $(t_0,y_0)\in D'$ *and* $L_G\neq 0$, *we also have:*

(3) $\lim_{(t,y)\to(t_0,y_0)}(F/G)(t,y) = L_F/L_G$.

2.2. Continuity

The most basic property we might wish for a function $f : D \to \mathbb{R}$ to have is that it is *continuous*. Here is the definition:

Definition 2.2.1. Suppose $f : D \to \mathbb{R}$ is a function with nice domain $D \subseteq \mathbb{R}$. We say that f is **continuous** if for every $\alpha \in D$,

$$\lim_{x\to\alpha} f(x) \;=\; f(\alpha).$$

Example 2.2.2. Here are some continuous functions:

(1) Every constant function $x \mapsto c\colon \mathbb{R} \to \mathbb{R}$ (where $c \in \mathbb{R}$) is continuous.
(2) The identity function $x \mapsto x\colon \mathbb{R} \to \mathbb{R}$ is continuous.
(3) The absolute value function $x \mapsto |x| := \sqrt{x^2}\colon \mathbb{R} \to \mathbb{R}$ is continuous.
(4) The square root function $x \mapsto \sqrt{x}\colon [0,+\infty) \to \mathbb{R}$ is also continuous.

The following shows how continuity is preserved under the basic arithmetic operations:

Proposition 2.2.3. *Suppose* $f, g\colon D \to \mathbb{R}$ *are continuous functions on a nice domain* D. *Then the following functions are also continuous on* D*:*

(1) $f + g\colon D \to \mathbb{R}$,
(2) $f \cdot g\colon D \to \mathbb{R}$

Furthermore, define the set

$$D' \;:=\; \big\{x\in D : g(x)\neq 0\big\}$$

and assume that D' *is nice (for us it always will be). Then*

(3) $f/g\colon D' \to \mathbb{R}$ *is continuous.*

The following tells us that continuity is preserved when you compose two composable continuous functions:

Proposition 2.2.4 (Composition and continuity)**.** *Suppose* $f\colon D \to \mathbb{R}$ *is continuous with nice domain* D *and* $g\colon E \to \mathbb{R}$ *is continuous with nice domain* E *such that* $f(D) \subseteq E$. *Then* $g \circ f\colon D \to \mathbb{R}$ *is continuous.*

Combining Example 2.2.2(3) with Proposition 2.2.4 gives us:

Corollary 2.2.5. *If* $f\colon D \to \mathbb{R}$ *is continuous with nice domain* D, *then so is* $|f|\colon D \to \mathbb{R}$, *given by*

$$|f|(x) := |f(x)|, \quad \textit{for } x \in D.$$

The following is an important theorem about continuous functions:

Intermediate Value Theorem 2.2.6. *Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous, with $a < b \in \mathbb{R}$. Let y be a number strictly between $f(a)$ and $f(b)$, i.e.,*

$$f(a) < y < f(b) \quad \text{or} \quad f(b) < y < f(a).$$

Then there is $x_0 \in (a,b)$ such that $f(x_0) = y$.

The following lemma says that if a continuous function is nonzero at a point, then it must be nonzero on a neighborhood of that point:

Bump Lemma 2.2.7. *Suppose $f : I \to \mathbb{R}$ is continuous, $I \subseteq \mathbb{R}$ is an interval, and $t_0 \in I$ is such that $f(t_0) \neq 0$. Then there is $\alpha < t_0 < \beta$ such that for every $t \in (\alpha, \beta) \cap I$, $f(t) \neq 0$.*

Monotonicity and inverses. In this subsection, we discuss monotone functions, the existence of inverse functions, and when inverse functions are continuous.

Definition 2.2.8. Suppose $f : D \to \mathbb{R}$ is a function where $D \subseteq \mathbb{R}$ is a nice set. We say that f is

(1) **increasing** if for all $x, y \in D$, if $x \leq y$, then $f(x) \leq f(y)$,
(2) **strictly increasing** if for all $x, y \in D$, if $x < y$, then $f(x) < f(y)$,
(3) **decreasing** if for all $x, y \in D$, if $x \leq y$, then $f(x) \geq f(y)$,
(4) **strictly decreasing** if for all $x, y \in D$, if $x < y$, then $f(x) > f(y)$.

Furthermore, we say that f is **monotone** if it satisfies any of properties (1)-(4), and we say that f is **strictly monotone** if it satisfies property (2) or (4).

Definition 2.2.9. Suppose $f : D \to \mathbb{R}$ is an injective function (see Definition B.5.1), and $D \subseteq \mathbb{R}$ is a nice set. We define the **inverse function** of f to be the function $f^{-1} : \text{range}(f) \to \mathbb{R}$ defined by:

$$f^{-1}(y) \;=\; x \quad :\Longleftrightarrow \quad f(x) \;=\; y$$

for all $x \in D$ and $y \in \text{range}(f)$.

Strictly monotone functions are a big source of injective functions:

Theorem 2.2.10. *Suppose $f : D \to \mathbb{R}$ is a strictly monotone function and D is a nice set. Then f is injective and so it has an inverse function $f^{-1} : f(D) \to \mathbb{R}$. Moreover, if one of the following holds:*

(1) f is continuous, or
(2) D is an interval,

then f^{-1} is continuous and strictly monotone.

Multivariable functions. There is also a definition of what it means for a multivariable function to be continuous:

Definition 2.2.11. Suppose $F : D \to \mathbb{R}$ is a two-variable function with domain $D \subseteq \mathbb{R}^2$ a nice subset of the ty-plane. We say that F is **continuous** (on D) if for every point $(t_0, y_0) \in D$, we have:

$$\lim_{(t,y)\to(t_0,y_0)} F(t,y) \;=\; F(t_0,y_0).$$

Most of the multivariable functions we will consider will be continuous, and their continuity can be determined by using the following rules, as well as the continuity of the underlying single-variable functions:

Continuity Laws for Multivariable Functions 2.2.12. *Suppose $F, G : D \to \mathbb{R}$ are continuous functions with domain $D \subseteq \mathbb{R}^2$ a nice subset of the ty-plane. Then:*

1. *(Projection functions) the functions $f(t, y) = t$ and $g(t, y) = y$ are continuous, as functions $f, g : D \to \mathbb{R}$.*
2. *(Linearity) Given arbitrary $\alpha, \beta \in \mathbb{R}$, the function:*
$$\alpha F + \beta G : D \to \mathbb{R}$$
is also continuous.
3. *(Products) The function:*
$$F \cdot G : D \to \mathbb{R}$$
is also continuous.
4. *(Quotients) Define the set*
$$D' \ := \ \big\{(t, y) \in D : G(t, y) \neq 0\big\}$$
Then the function:
$$\frac{F}{G} : D' \to \mathbb{R}$$
is also continuous.
5. *(Compositions) Suppose $f : E \to \mathbb{R}$ is a continuous one-variable function where $E \subseteq \mathbb{R}$ is a nice domain. Furthermore, suppose $F(D) \subseteq E$. Then the composition:*
$$f \circ F : D \to \mathbb{R}$$
is also a continuous function.

2.3. Differentiation

In this section $D \subseteq \mathbb{R}$ is a nice set. Given a function $f : D \to \mathbb{R}$, if it is differentiable at a point in its domain, then that means the function f can be approximated suspiciously well by a linear tangent line at that point. The following proposition gives three equivalent ways of saying exactly this:

Proposition 2.3.1. *Suppose $f\colon D \to \mathbb{R}$ is a function and $\alpha \in D$. The following are equivalent:*

1. *(Standard definition) The limit*
$$\lim_{x \to \alpha} \frac{f(x) - f(\alpha)}{x - \alpha} \ = \ \ell$$
exists and is finite (i.e., $\ell \in \mathbb{R}$).
2. *(Taylor definition) There exists a number $d \in \mathbb{R}$ and a function $R\colon D \to \mathbb{R}$ such that*
$$f(x) \ = \ f(\alpha) + d(x - \alpha) + R(x) \quad \text{and} \quad \lim_{x \to \alpha} \frac{R(x)}{x - \alpha} \ = \ 0.$$
3. *(Carathéodory definition) There exists a function $q\colon D \to \mathbb{R}$ which is continuous at α such that*
$$f(x) \ = \ f(\alpha) + q(x)(x - \alpha).$$

Furthermore, if any (equivalently all) of (1), (2), and (3) holds, then

4. *$\ell = d = q(\alpha)$, and*
5. *f is continuous at α.*

Definition 2.3.2. We say that function $f : D \to \mathbb{R}$ is **differentiable on** D, if for every $\alpha \in D$, the equivalent conditions of Proposition 2.3.1 hold. In this case, we define the **derivative of** f **at** α to be

$$f'(\alpha) \ := \ \lim_{x\to\alpha} \frac{f(x)-f(\alpha)}{x-\alpha}.$$

In this class, since we will be working with special elementary functions and not arbitrary differentiable functions, we generally will not have to use the formal definition when computing derivatives. In general we will be able to compute all relevant derivatives by employing the following rules as well as the known formulas (see Appendix A) for the derivatives of the functions we care about.

Example 2.3.3. (1) Constant functions are differentiable with derivative 0.

(2) Let $f\colon \mathbb{R} \to \mathbb{R}$ be such that $f(x) = x^n$. Then f is differentiable, and and for every $\alpha \in \mathbb{R}$ $f'(\alpha) = n\alpha^{n-1}$. To see this, note by The Difference of Powers Formula,

$$f(x)-f(\alpha) \ = \ x^n-\alpha^n \ = \ (x-\alpha)\cdot(x^{n-1}+\alpha x^{n-2}+\alpha^2x^{n-3}+\cdots+\alpha^{n-2}x+\alpha^{n-1}),$$

thus for $x \neq \alpha$, we have

$$\frac{f(x)-f(\alpha)}{x-\alpha} \ = \ x^{n-1}+\alpha x^{n-2}+\alpha^2x^{n-3}+\cdots+\alpha^{n-2}x+\alpha^{n-1},$$

and so

$$\lim_{x\to\alpha} \frac{f(x)-f(\alpha)}{x-\alpha} \ = \ n\cdot\alpha^{n-1}.$$

The following rules show how computing the derivative interacts with the basic arithmetic operations:

Proposition 2.3.4. *Suppose* $f, g\colon D \to \mathbb{R}$ *are differentiable on* D*. Then*

$$f+g, f\cdot g\colon D \to \mathbb{R}$$

are differentiable on D*, and for every* $\alpha \in D$

(1) $(f+g)'(\alpha) = f'(\alpha)+g'(\alpha)$,

(2) (product rule) $(f\cdot g)'(a) = f(a)g'(a)+f'(a)g(a)$,

Furthermore, with $D' := \{x \in D : g(x) \neq 0\} \subseteq D$*, if* D' *is nice, then the function*

$$\frac{f}{g} : D' \to \mathbb{R}$$

is differentiable and

(3) (quotient rule) for every $\alpha \in D'$

$$\left(\frac{f}{g}\right)'(\alpha) \ = \ \frac{g(\alpha)f'(\alpha)-f(\alpha)g'(\alpha)}{g^2(\alpha)}$$

Remark 2.3.5. An immediate consequence of Proposition 2.3.4(1) and (2) is that if we have constants $c, d \in \mathbb{R}$ and differentiable functions $f, g\colon D \to \mathbb{R}$, then

$$(cf+dg)' \ = \ cf'+dg'.$$

In linear algebra terms, differentiation is $\mathbb{R}$-*linear* (i.e., it is a linear transformation on the $\mathbb{R}$-vector space of differentiable functions $D \to \mathbb{R}$).

Differentiation also behaves well with *composition* of differentiable functions:

Chain Rule 2.3.6. *Suppose $f\colon D \to \mathbb{R}$, $g\colon E \to \mathbb{R}$ are differentiable functions such that $f(D) \subseteq E$. Then $g \circ f : D \to \mathbb{R}$ is differentiable, and for every $\alpha \in D$*

$$(g \circ f)'(\alpha) \;=\; g'\big(f(\alpha)\big) \cdot f'(\alpha).$$

In theory, you should be able capable of computing the derivative of any elementary function provided you know the rules 2.3.4 and 2.3.6 as well as the formulas for the derivatives of the primitive functions of interest given in Appendix A. Of course, this should not be news to you.

The following is a very useful consequence of the so-called *Mean Value Theorem for Derivatives*. Note that Corollary 2.3.7 and Identity Criterion 2.3.8 are only true when the domain is an interval.

Corollary 2.3.7. *Suppose D is an interval and $f\colon D \to \mathbb{R}$ is differentiable. Then f is a constant function iff $f'(x) = 0$ for all $x \in I$.*

A common question we might ask when it comes to *uniqueness* of solutions of ODEs is: when are two functions $f, g\colon I \to \mathbb{R}$ the same? If f and g are differentiable (which pretty much all of our functions will be), the following makes this question easier to answer:

Identity Criterion 2.3.8. *Suppose D is an interval and $f, g\colon D \to \mathbb{R}$ are differentiable such that $f'(\alpha) = g'(\alpha)$ for every $\alpha \in D$. Then there exists a constant $C \in \mathbb{R}$ such that $f(x) = g(x) + C$ for all $x \in D$. Furthermore, if there is a point $x_0 \in D$ such that $f(x_0) = g(x_0)$, then $f(x) = g(x)$ for all $x \in D$.*

Proof. The function $f - g\colon D \to \mathbb{R}$ is differentiable by Proposition 2.3.4, and $(f-g)'(x) = f'(x) - g'(x) = 0$ for all $x \in D$. By Corollary 2.3.7, there is a constant $C \in \mathbb{R}$ such that $(f-g)(x) = C$ for all $x \in D$, i.e., $f(x) = g(x) + C$ for all $x \in D$.

Now, suppose there is $x_0 \in D$ such that $f(x_0) = g(x_0)$. Then also $f(x_0) = g(x_0) + C$, so we can conclude that $C = 0$. Thus $f(x) = g(x)$ for all $x \in D$. □

Inverse functions and monotonicity. Sometimes differentiable functions are also invertible. In this subsection we talk about the differentiability of the inverse function.

Theorem 2.3.9. *Assume $f : D \to \mathbb{R}$ is a differentiable injective function and $D \subseteq \mathbb{R}$ is a nice set. Define $I := f(D)$ and*

$$I' \;:=\; \{y \in I : f'\big(f^{-1}(y)\big) \neq 0\}$$

The the function $f^{-1} : I' \to \mathbb{R}$ is differentiable, and for every $y_0 \in I'$ we have

$$(f^{-1})'(y_0) \;=\; \frac{1}{f'\big(f^{-1}(y_0)\big)}$$

We can also use derivatives to check for monotonicity, which enable us show that a function is invertible.

Theorem 2.3.10. *Suppose $f : I \to \mathbb{R}$ is a differentiable function on an interval I. Then:*

(1) f is increasing if $f'(x) \geq 0$ for all $x \in I$,
(2) f is strictly increasing if $f'(x) > 0$ for all $x \in I$,
(3) f is decreasing if $f'(x) \leq 0$ for all $x \in I$, and
(4) f is strictly decreasing if $f'(x) < 0$ for all $x \in I$.

Multivariable functions. A full exploration of multivariable calculus (differentiation and integration) requires a course like Math32A or Math131B. For our purposes, we will need to know a few things about partial derivatives:

Definition 2.3.11. Suppose $F : D \to \mathbb{R}$ is a function with nice domain $D \subseteq \mathbb{R}^2$ (so $F = F(t, y)$ is a two-variable function). Let $(t_0, y_0) \in D$ be a fixed point. We define the **partial derivative of F with respect to t at** (t_0, y_0) to be the following limit, if it exists and is finite:

$$\frac{\partial F}{\partial t}(t_0, y_0) := \lim_{t \to 0} \frac{F(t_0 + t, y_0) - F(t_0, y_0)}{t}$$

and we define the **partial derivative of F with respect to y at** (t_0, y_0) to be the following limit, if it exists and is finite:

$$\frac{\partial F}{\partial y}(t_0, y_0) := \lim_{y \to 0} \frac{F(t_0, y_0 + y) - F(t_0, y_0)}{y}$$

In practice, a partial derivative is the same thing as a single-variable derivative where you treat the other variable as a constant. In particular, all of the rules from the preceding subsection apply to partial derivatives when you view them this way (product rule, chain rule, etc.).

Definition 2.3.12. Suppose $D \subseteq \mathbb{R}^2$ is a nice subset of $\mathbb{R}^2$, and $F : D \to \mathbb{R}$ is a two-variable function. We say that:

(1) F **has first-order partial derivatives** if at every point $(t_0, y_0) \in D$, the partial derivatives

$$\frac{\partial F}{\partial t}(t_0, y_0) \quad \text{and} \quad \frac{\partial F}{\partial y}(t_0, y_0)$$

exist and are finite;

(2) F **has second-order partial derivatives** if:
   (i) F has first-order partial derivatives, and
   (ii) the functions $\frac{\partial F}{\partial t}, \frac{\partial F}{\partial y} : D \to \mathbb{R}$ also have first order partial derivatives.

(3) F **has continuous second-order partial derivatives** if:
   (a) F has second-order derivatives, and
   (b) each of the functions:

$$\frac{\partial^2 F}{\partial t^2}, \frac{\partial^2 F}{\partial t \partial y}, \frac{\partial^2 F}{\partial y \partial t}, \frac{\partial^2 F}{\partial y^2} : D \to \mathbb{R}$$

   are continuous.

In general, all of the two-variable functions we'll consider have continuous partial derivatives of all orders, including first and second order, at least wherever they are defined. In this case, the following theorem tells us that the "mixed" second order partial derivatives are the same. This will be useful for getting a checkable criterion for exactness in Section 3.5.

Clairaut-Schwarz Theorem 2.3.13 (Equality of mixed partial derivatives)**.** *Suppose $F : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^2$ is a nice subset of the plane $\mathbb{R}^2$ has continuous second-order partial derivatives, i.e.,*

$$\frac{\partial^2 F}{\partial t^2}, \quad \frac{\partial^2 F}{\partial y^2}, \quad \frac{\partial^2 F}{\partial t \partial y}, \quad \frac{\partial^2 F}{\partial y \partial t}$$

all exist and are continuous (the functions we'll deal with always satisfy this property). Then for all $(t_0, y_0) \in D$*:*

$$\frac{\partial^2 F}{\partial t \partial y}(t_0, y_0) \;=\; \frac{\partial^2 F}{\partial y \partial t}(t_0, y_0)$$

2.4. Integration

Definite integrals. When it comes to integration, the most fundamental notion is to define the following: given a function $f : [a, b] \to \mathbb{R}$, what does it mean for the function f to be *integrable* on $[a, b]$ and how do you define $\int_a^b f(t)\,dt$ if this integral is to exist? We will not dive into this question and instead assume you have a working understanding of what this means to you. In particular, we define:

Definition 2.4.1. Suppose $a < b \in \mathbb{R}$. We say that the function $f : [a, b] \to \mathbb{R}$ is **integrable** if the definite integral

$$\int_a^b f(t)\,dt$$

exists and is finite (i.e., it equals a real number from $\mathbb{R}$). If $f : [a, b] \to \mathbb{R}$ is integrable, then we also define:

$$\int_b^a f(t)\,dt \;:=\; -\int_a^b f(t)\,dt$$

Given any function $g : D \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, we define:

$$\int_\alpha^\alpha g(t)\,dt \;:=\; 0$$

Here are some basic facts about what types of functions are integrable:

Fact 2.4.2. Suppose $f : [a, b] \to \mathbb{R}$ is a function. Then:

(1) if f is continuous, then f is integrable,
(2) if f is piecewise continuous, then f is integrable,
(3) if f is integrable and $\tilde{f} : [a, b] \to \mathbb{R}$ is a function such that the set:

$$\big\{x \in [a, b] : f(x) \neq \tilde{f}(x)\big\}$$

is finite, then $\tilde{f}$ is also integrable and

$$\int_a^b f(t)\,dt \;=\; \int_a^b \tilde{f}(t)\,dt$$

Fact 2.4.2 tells us that basically every function $f : [a, b] \to \mathbb{R}$ we come across in this class will be integrable. Furthermore, 2.4.2(3) tells us that as far as computing integrals are concerned, we can safely change finitely many values of the function and still arrive at the same answer (for instance, if you are integrating a step function and you're not sure about the values at the endpoints).

The following law for computing definite integrals is used all the time:

Lemma 2.4.3 (Linearity of Integration)**.** *Let* $f, g\colon [a, b] \to \mathbb{R}$ *be integrable functions, and let* $\alpha \in \mathbb{R}$*. Then*

(1) $\alpha f\colon [a, b] \to \mathbb{R}$ *is integrable, and* $\int_a^b \alpha f(t)\,dt = \alpha \int_a^b f(t)\,dt$*,*
(2) $f + g\colon [a, b] \to \mathbb{R}$ *is integrable, and* $\int_a^b (f + g)(t)\,dt = \int_a^b f(t)\,dt + \int_a^b g(t)\,dt$*.*

The following is also very useful, especially if the behavior of a function changes on different intervals:

Lemma 2.4.4 (Additivity over intervals)**.** *Suppose $f\colon [a,b] \to \mathbb{R}$ is a function and $c \in (a,b)$. Then f is integrable on $[a,b]$ iff f is integrable on $[a,c]$ and $[c,b]$. In this case, we have*

$$\int_a^b f(t)\,dt \;=\; \int_a^c f(t)\,dt + \int_c^b f(t)\,dt.$$

The following two theorems tell us that *integration* and *differentiation* are inverse operations, which is what makes integration so useful when it comes to solving differential equations. First a definition:

Definition 2.4.5. Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$. A function $F : D \to \mathbb{R}$ is called an **antiderivative** of f if:

(i) F is differentiable, and
(ii) for every $t \in D$, $F'(t) = f(t)$.

The so-called *first fundamental theorem of calculus* provides us a method of computing the exact value of the definite integral of a function provided we have available to us an antiderivative of that function:

First Fundamental Theorem of Calculus 2.4.6. *Suppose $f : [a,b] \to \mathbb{R}$ is a continuous function on $[a,b]$ and differentiable on (a,b). Then:*

$$\int_a^b f'(t)\,dt \;=\; f(a) - f(b).$$

The so-called *second fundamental theorem of calculus* provides us a method of using definite integrals to construct an antiderivative of a continuous function:

Second Fundamental Theorem of Calculus 2.4.7. *Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$, and fix $t_0 \in D$. Let $I \subseteq D$ be the largest interval such that $t_0 \in I$. Consider the function $F : I \to \mathbb{R}$ defined by*

$$F(t) \;:=\; \int_{t_0}^t f(s)\,ds$$

for every $t \in I$. Then

(1) F is differentiable on I, and
(2) $F'(t) = f(t)$ for every $t \in I$, i.e., F is an antiderivative of f on the interval I.

Indefinite integrals. When we later determine the general solution of a differential equation, we need to be able to find (and parametrize) *all* solutions of the differential equation, not just a particular one. In terms of antiderivatives, this means we need to be able to find (and parametrize) *all* antiderivatives of a particular function, not just one antiderivative. This is taken care of by the notion of *indefinite integral*:

Definition 2.4.8. Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$. The **indefinite integral** of f is an infinite family of functions:

$$F(t;C) \;=\; F(t) + C$$

where $C \in \mathbb{R}$ and $F : D \to \mathbb{R}$ is a particular antiderivative of f. This situation is often denoted by writing:

$$\int f(t)\,dt \;=\; F(t) + C.$$

Remark 2.4.9. Technically speaking, the indefinite integral of f really should be the family of *all* antiderivatives of f. In particular, each so-called *connected component* of the domain of f requires its own constant of integration. For instance, for the function $f(t) = 1/t$ viewed as a function $(-\infty, 0) \cup (0, +\infty) \to \mathbb{R}$, the indefinite integral really should be:

$$\int \frac{dt}{t} \;=\; \begin{cases} \ln(t) + C_1 & \text{if } t > 0 \\ \ln(-t) + C_2 & \text{if } t < 0 \end{cases}$$

where $C_1, C_2 \in \mathbb{R}$ could be the same number, or could be different. Simply writing:

$$\int \frac{dt}{t} \;=\; \ln|t| + C$$

does not actually give us every possible antiderivative of $1/t$ on the domain $(-\infty, 0) \cup (0, +\infty)$ because it requires us to use the same constant of integration on both "connected components" $(-\infty, 0)$ and $(0, +\infty)$. This is a very minor issue which we are happy to ignore since the particular solutions to initial value problems (which we hope to be unique) will have intervals as their domain.

We also have the second fundamental theorem of calculus for indefinite integrals:

Second Fundamental Theorem of Calculus 2.4.10 (Indefinite version)**.** *Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$. Then*

$$\frac{d}{dt} \int f(t)\,dt \;=\; f(t).$$

Theorem 2.4.10 is to be interpreted as: for every antiderivative $F(t) + C$ of $f(t)$,

$$\frac{d}{dt}\big(F(t) + C\big) \;=\; f(t).$$

CHAPTER 3

First-order differential equations

3.1. Implicit differential equations

In this course we will be primarily concerned with first-order differential equations, as well as higher-order *linear* differential equations. This begs the question:

What is a differential equation and what is the order of a differential equation?

We will answer this question by first giving a very general definition of *differential equation* which will encompass nearly all differential equations we will encounter in this Chapter and in Chapter 4:

Definition 3.1.1. An **implicit differential equation (of order r)** is an equation which can be written in the form

$$F(t, y, y', y'', \dots, y^{(r)}) \;=\; 0 \tag{$\dagger$}$$

where F is a real-valued function of $r+2$ variables. The **order** is the order r of the highest derivative $y^{(r)}$ of y which appears in the equation.

A **solution** to ($\dagger$) is a function $y : I \to \mathbb{R}$ (where $I \subseteq \mathbb{R}$ is an interval) which is differentiable at least r times such that

$$F\big(t, y(t), y'(t), \dots, y^{(r)}(t)\big) \;=\; 0 \quad \text{for every } t \in I,$$

i.e., for every $t \in I$, when you plug $t, y(t), y'(t), \dots, y^{(r)}(t)$ into the function F the output is zero.

We now give some examples of implicit differential equations and some of their solutions, in increasing order of order.

Zeroth order. Here is an implicit differential equation of order 0:

$$y^5 + 2y^4 + 3y^3 + 4y^2 + 5y + 6 \;=\; 0 \tag{3.1}$$

Given a solution $\alpha \in \mathbb{R}$ of the polynomial equation

$$X^5 + 2X^4 + 3X^3 + 4X^2 + 5X + 6 \;=\; 0,$$

the function $y : \mathbb{R} \to \mathbb{R}$ defined by $y(t) := \alpha$ for all $t \in \mathbb{R}$ (i.e., the function with constant value α) is a solution of (3.1). This example should convince you that the subject of differential equations already encompasses all of one- and two-variable polynomial equations. In particular, we shouldn't get our hopes up that we will be able to solve too many higher-order differential equations in general.

First order. We will give two examples of a first-order differential equation. The first one takes full advantage of the *implicit* part of the definition:

Example 3.1.2 (Clairaut)**.** The differential equation:

$$y - ty' + \exp y' = 0 \tag{3.2}$$

Every solution $y : \mathbb{R} \to \mathbb{R}$ of (3.2) has the form

$$y(t) = Ct + \exp C$$

where $C \in \mathbb{R}$ is some fixed constant. Note that even though (3.2) is complicated, it is actually pretty easy to check that the given solution is actually correct. Indeed, first compute the derivative of y:

$$y'(t) = C$$

and then plug $t, y(t), y'(t)$ into (3.2) and notice that everything cancels out:

$$y(t) - ty'(t) + \exp y'(t) = Ct + \exp C - tC + \exp C = 0.$$

This illustrates another important lesson:

Checking that a given function is/is not a solution to a differential equation is usually easy, even if the given differential equation is hard/impossible.

Indeed, it is simply a matter of computing r derivatives and then plugging them into the equation and seeing if everything cancels out. Of course, we will be more interested in solving differential equations than checking whether a candidate solution is correct or not. However, it is reassuring to know that at least one direction of the process is fairly easy.

The next differential equation is a more typical example of a differential equation which we will study:

Example 3.1.3 (Logistic equation)**.** Let $b, c > 0$ be fixed positive constants. Then the **logistic equation** is the differential equation:

$$y' - y(b - cy) = 0$$

For every nonzero constant $C \in \mathbb{R} \setminus \{0\}$ we have a solution $y : \mathbb{R} \to \mathbb{R}$ defined by:

$$y(t) = \frac{b}{c} \cdot \frac{1}{1 + C\exp(-bt)}$$

Furthermore, the constant functions $y = 0$ and $y = b/c$ are also solutions. (Exercise: check this!) We will study the logistic equation in more detail later, including how to derive these solutions.

Second order. Here is a typical example of a second-order differential equation we will study:

$$y'' - 3y' + 2y = 0 \tag{3.3}$$

Every solution $y : \mathbb{R} \to \mathbb{R}$ of (3.3) is of the form:

$$y(t) = C_1 \exp 2t + C_2 \exp t$$

where $C_1, C_2 \in \mathbb{R}$ are arbitrary constants. Generally speaking, for second-order differential equations there will be two constants of integration we need to find. This reflects the fact that the equation involves a first and second derivative (so

somewhere we are doing two integrals, each one with its own constant of integration). Equation (3.3) is an example of a *second-order linear differential equation with constant coefficients*, which will be one of the main equations of interest in Chapter 4.

3.2. Differential equations in normal form

Definition 3.1.1 casts a very wide net. In general most differential equations we will encounter can be put into a slightly simpler form: *normal form*.

Definition 3.2.1. A **differential equation of order r in normal form** (or an **explicit differential equation of order** r) is a differential equation which can be written in the form

$$(\dagger) \qquad y^{(r)} \;=\; F(t, y, y', y'', \dots, y^{(r-1)})$$

where F is a real-valued function of $r+1$ variables. A **solution** of (†) is a function $y : I \to \mathbb{R}$ (where $I \subseteq \mathbb{R}$ is an interval) which is at least r times differentiable, such that for every $t \in I$:

$$y^{(r)}(t) \;=\; F\big(t, y(t), y'(t), \dots, y^{(r-1)}(t)\big)$$

In other words, an implicit differential equation of order r can be put into normal form if it is possible to solve for the highest derivative $y^{(r)}$ in terms of the lower derivative $y, y', \dots, y^{(r-1)}$ and t.

Example 3.2.2.

(1) A zeroth-order differential equation in normal form is an equation of the form:

$$y \;=\; F(t)$$

Clearly, the function $y(t) := F(t)$ is a solution. We will never be interested in explicit zeroth-order differential equations.

(2) A first-order differential equation in normal form is an equation of the form:

$$y' \;=\; F(t, y)$$

The logistic equation from Example 3.1.3 can be put into normal form:

$$y' \;=\; y(b - cy)$$

It is not clear whether the equation from Example 3.1.2

$$y - ty' + \exp y' \;=\; 0$$

can be put into normal form since this would involve solving for y'. In general, for the equations we deal with there will be no issue with rewriting them in normal form.

(3) A second-order differential equation in normal form is an equation of the form:

$$y'' \;=\; F(t, y, y').$$

Equation (3.3) can be written in normal form:

$$y'' \;=\; 3y' - 2y$$

This concludes our discussion of general-order differential equations. For the rest of the chapter we will focus on first-order differential equations in normal form.

Explicit first-order differential equations. Recall that an explicit first-order differential equation is an equation which can be written in the form:

$$y' = F(t, y) \tag{3.4}$$

where F is a real-valued function of two variables. A **solution** to (3.4) is a differentiable function $y : I \to \mathbb{R}$ ($I \subseteq \mathbb{R}$ is an interval) such that for all $t \in I$,

$$y'(t) = F\big(t, y(t)\big)$$

Solutions are also referred to as **integral curves** or **solution curves**, especially when we want to emphasize the geometric properties of the solution.

We will often be interested in obtaining a specific solutions which passes through a given point $(t_0, y(t_0))$. The best way to do this is to first find *all* solutions of the differential equation, and then find the particular solution we are interested in.

Definition 3.2.3. The **general solution** of (3.4) is a family[1] of functions $y(t; C)$ which depends on a parameter $C \in \mathbb{R}$ such that:

(1) for every valid parameter C_0, the function $y(t; C_0)$ is a solution of (3.4), and
(2) every solution of (3.4) is of the form $y(t; C_1)$ for some valid parameter C_1.

A **particular solution** is a function of the form $y(t) = y(t; C_0)$ for some fixed value C_0.

Example 3.2.4. Consider the differential equation

$$y' = t \tag{3.5}$$

We wish to find the general solution to (3.5). Integrating both sides, we find that

$$y(t) = \frac{1}{2}t^2 + C$$

for some constant of integration $C \in \mathbb{R}$. We claim that the general solution is

$$y(t; C) = \frac{1}{2}t^2 + C$$

where C can be any real number. Indeed, for every specific $C_0 \in \mathbb{R}$, the function $y(t) = \frac{1}{2}t^2 + C_0$ is a solution. Furthermore, if $\bar{y}(t)$ is also a solution, then $\bar{y}'(t) = t$, and thus

$$\big(\bar{y}(t) - y(t; 0)\big)' = \big(\bar{y}(t) - \frac{1}{2}t^2\big)' = t - t = 0$$

which shows that $\bar{y}(t)$ and $y(t; 0)$ differ by a constant. Thus there exists $C_1 \in \mathbb{R}$ such that $\bar{y}(t) = y(t; C_1)$. We conclude that $y(t; C)$ is the general solution of (3.5). Here are some particular solutions:

$$y(t) = y(t; 3) = \frac{1}{2}t^2 + 3$$
$$y(t) = y(t; -10) = \frac{1}{2}t^2 - 10.$$

The problem of finding a specific particular solution will be formulated as an *initial value problem*:

[1]The notation $y(t; C)$ is meant to suggest that the function $y(t)$ depends also on the parameter C. Each time you choose a specific value C_0 for C, then you get a particular solution $y(t) := y(t; C_0)$.

Definition 3.2.5. An **initial value problem** is a pair of two conditions:

(i) a differential equation:
$$y' = F(t, y)$$

(ii) a specific point which the solution must pass through:
$$y(t_0) = y_0,$$
where $(t_0, y_0) \in \mathbb{R}^2$. This is called the **initial condition**.

Example 3.2.6. We wish to solve the following initial value problem:

(i) $y' = t$
(ii) $y(3) = 7$

We have already found that the general solution to (i) is
$$y(t; C) = \frac{1}{2}t^2 + C$$
We will use (ii) to solve for the exact value of C:
$$y(3) = 7 = \frac{1}{2} \cdot 3^2 + C$$
and so
$$C = 7 - \frac{9}{2} = \frac{5}{2}.$$
We conclude that the solution to the above initial value problem is:
$$y(t) = y(t; 5/2) = \frac{1}{2}t^2 + \frac{5}{2}.$$

Direction fields. One of the remarkable features of explicit first-order differential equations is that, even if some of them might be difficult to solve, it is usually pretty easy to make a rough sketch of the general solutions. This is because the equation
$$y' = F(t, y)$$
tells us what the derivative of the solution needs to be at each point (t, y) in the plane. We make this precise with the notion of a *direction field*.

Definition 3.2.7. A **direction field** for the equation
$$y' = F(t, y)$$
is a plot where at each point (t_0, y_0) you draw a tiny line segment with slope $F(t_0, y_0)$.

Of course in practice when you (or a computer) draw a direction field, you can't possibly draw such a line segment at every point in the plane (since there are infinitely many such points). Instead you draw enough tiny line segments (say, at integer or half-integer coordinates) in order to get a sense of the general behavior of the direction field. Once you have an accurate direction field, you can sketch an approximation of a solution by "following the direction of the direction field".

Example 3.2.8. Consider the logistic equation
$$y' = y(3 - y) \tag{3.6}$$
In Figure 3.1 we plot the direction field for (3.6). We also include four solution curves corresponding to four different initial conditions.

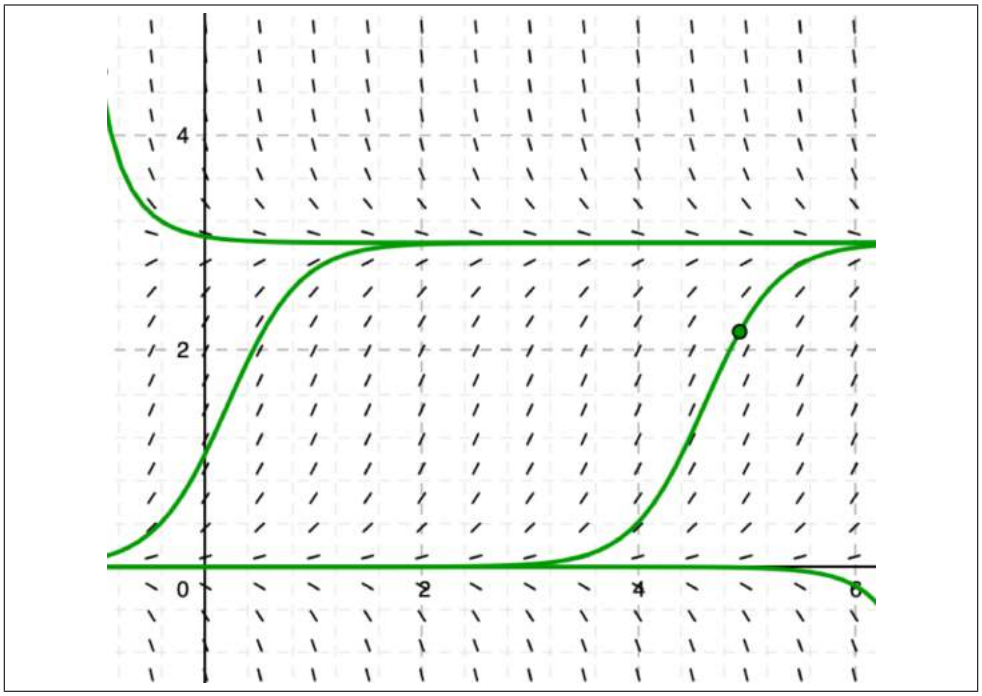


FIGURE 3.1. Direction field for the logistic equation $y' = y(3-y)$ and several solution curves.

We make the following observations:

(1) At each point (t_0, y_0), the slope only depends on y_0. This is because $y(3-y)$ only depends on y and not on t.
(2) This suggests that if $y(t)$ is a solution to (3.6), then so is $y(t+C)$ for any constant C.
(3) The direction field suggests that the constant functions
$$y(t) \;=\; 0 \quad \text{and} \quad y(t) \;=\; 3$$
are both solutions to (3.6). This is indeed the case, as can be easily verified.
(4) There are many other non-constant solutions as well, we will learn how to solve for them in Section 3.5.

Of course, by merely plotting a direction field and sketching a solution curve, you are not actually solving the differential equation yet. However, this procedure provides valuable insight into the nature of the solutions which can be very fruitful. In some sense, this is the starting point for the *qualitative study of differential equations*.

3.3. First-order linear differential equations

We now arrive at the first family of differential equations which we will study in detail, the so-called *first-order linear differential equations*.

Definition 3.3.1. A **first-order linear differential equation** is a differential equation which can be written in the form:
$$y' + f(t)y \;=\; g(t)$$

where f, g are real-valued functions of the variable t. The function $f(t)$ and $g(t)$ are called[2] the **coefficient functions**.

As we shall see, solving a first-order linear differential equation really boils down to performing an integration. We will work up to the general case (where both $f(t)$ and $g(t)$ are nonzero functions) in several steps.

Direct integration. Consider first the case where $f(t) = 0$ for all t. We call the resulting differential equation:

$$y' \;=\; g(t)$$

a **direct integration** differential equation. This is because you can directly solve this differential equation by integrating g and, if need be, solving for C with the initial condition. Here is an example:

Example 3.3.2. Consider the initial value problem:

(i) $y' = \sqrt{t}$,
(ii) $y(4) = 6$.

Integrating the differential equation we obtain

$$y(t) \;=\; 2/3t^{3/2} + C.$$

Using the initial condition we get

$$y(4) \;=\; 6 \;=\; 2/3(4)^{3/2} + C$$

and so $C = 6 - 16/3 = 2/3$. So the solution to the above initial value problem is

$$y(t) \;=\; 2/3t^{3/2} + 2/3.$$

In Figure 3.2 we plot the corresponding solution curve together with the direction field. Notice that the solution exists on the interval $[0, +\infty)$, and this is the possible interval on which the solution can exist and remain a solution because $g(t) = \sqrt{t}$ is only defined on $[0, +\infty)$.

We also remark that in Figure 3.2 we see that the direction field only depends on t and not on y. This observation allows us to guess (if we didn't know it already) that any two solutions of (i) differ by a vertical translation (i.e., adding a constant). This indeed is also the case for general *direct integration* differential equations.

[2]sometimes just $f(t)$ is called the **coefficient function** and $g(t)$ is called the **forcing function**.

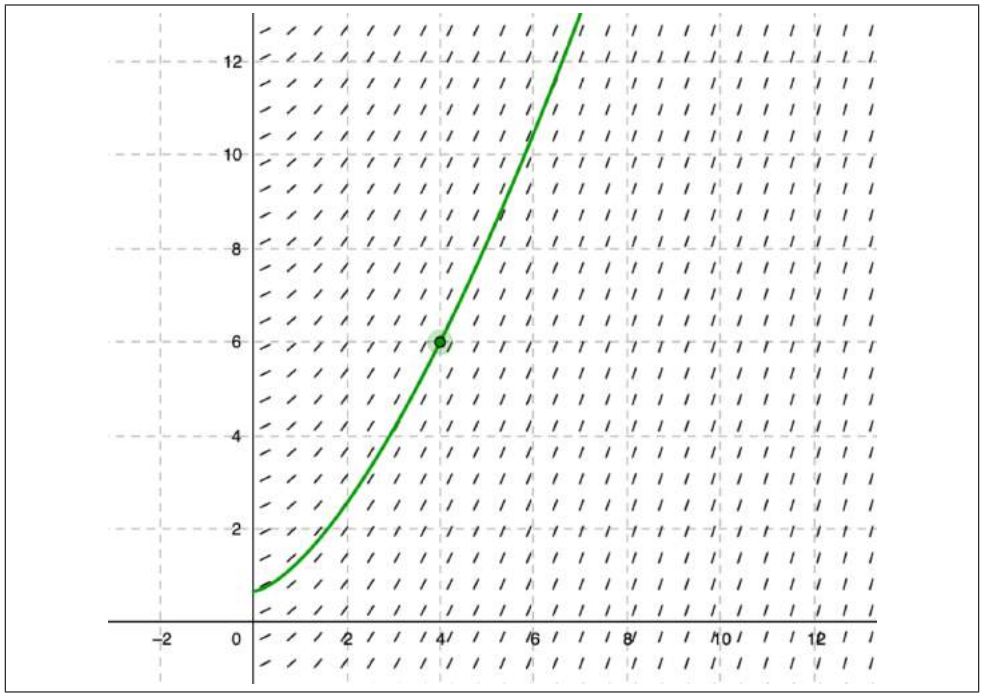


FIGURE 3.2. Direction field for the equation $y' = \sqrt{t}$ and the solution curve passing through the point $(4, 6)$.

Theorem 3.3.3 (Direct Integration)**.** *Suppose $g : D \to \mathbb{R}$ is a continuous function with nice domain $D \subseteq \mathbb{R}$. Consider the differential equation:*

(i) $y' = g(t)$

(1) The general solution of (i) is given by

$$y(t) \;=\; y(t;C) \;=\; \int g(t)\,dt + C$$

Furthermore, suppose we are also given an initial condition

(ii) $y(t_0) = y_0$*, where* $t_0 \in D$ *and* $y_0 \in \mathbb{R}$*.*

(2) Then the initial value problem (i)+(ii) has the unique solution:

$$y(t) \;=\; \int_{t_0}^{t} g(s)\,ds + y_0$$

(3) The ***interval of existence*** *of this solution (i.e., the largest interval containing t_0 for which this function remains a solution) is the largest interval $I \subseteq \mathbb{R}$ such that:*

(a) $t_0 \in I$*, and*

(b) $I \subseteq D$*.*

The homogeneous case. We next consider the case where $g(t)$ is the constant zero function and $f(t)$ is possibly nonzero.

Definition 3.3.4. A first-order linear differential equation is said to be **homogenous** if it is of the form:

$$y' + f(t)y \;=\; 0.$$

Solving the homogeneous case requires knowing a trick: multiplication by a so-called *integrating factor*. We illustrate this first with an example:

Example 3.3.5. Consider the homogeneous first-order linear differential equation:

$$y' + \frac{1}{t}y \;=\; 0 \tag{3.7}$$

Here we are regarding the coefficient function $1/t$ to have domain $(-\infty, 0) \cup (0, +\infty)$. First observe that if $\mu(t)$ is any function which is never zero, then the differential equation

$$\mu(t)\left(y' + \frac{1}{t}y\right) \;=\; 0$$

has the same solutions as equation (3.7). We will use the following choice of $\mu(t)$:

$$\mu(t) \;:=\; \exp\left(\int \frac{dt}{t}\right) \;=\; \exp \ln |t| \;=\; |t|$$

where the domain of $\mu(t)$ is also $(-\infty, 0) \cup (0, +\infty)$. Then we multiply the lefthand side of (3.7) by $\mu(t)$ to obtain:

$$|t|\left(y' + \frac{1}{t}y\right) \;=\; |t|y' + \operatorname{sgn}(t)y \;=\; \big(|t|y\big)' \;=\; 0.$$

In other words, multiplying through by the integrating factor $\mu(t)$ allows us to view the lefthand side as the derivative of a single function of t. Next we integrate both sides of

$$\big(|t|y\big)' \;=\; 0$$

to obtain

$$|t|y(t) \;=\; C,$$

or rather,

$$y(t) \;=\; \frac{C}{|t|}.$$

Here the function $y(t)$ also has domain $(-\infty, 0) \cup (0, +\infty)$.

Here is how to handle the general homogeneous case:

Theorem 3.3.6. *Suppose $f : D \to \mathbb{R}$ is a continuous function with nice domain $D \subseteq \mathbb{R}$ consider the differential equation:*

(i) $y' + f(t)y = 0$

(1) Define the ***integrating factor*** *to be the function $\mu : D \to \mathbb{R}$ given by:*

$$\mu(t) \;:=\; \exp\left(\int f(t)\,dt\right)$$

(here $\int f(t)\,dt$ can be any antiderivative of $f(t)$, the constant of integration does not matter). Then we can multiply (i) by μ to obtain:

$$\mu(t)\big(y' + f(t)y\big) \;=\; \big(\mu(t)y\big)' \;=\; 0.$$

(2) The general solution of (i) is given by:

$$y(t) \;=\; y(t; C) \;=\; \frac{C}{\mu(t)} \;=\; C\exp\left(-\int f(t)\,dt\right)$$

Furthermore, suppose we are also given an initial condition

(ii) $y(t_0) = y_0$, *where $t_0 \in D$ and $y_0 \in \mathbb{R}$.*

(3) Then the initial value problem (i)+(ii) has the unique solution:

$$y(t) \;=\; y_0 \exp\left(-\int_{t_0}^{t} f(s)\,ds\right) \;=\; \frac{y_0}{\mu(t)}$$

where $\mu(t) := \exp(\int_{t_0}^{t} f(s)\,ds)$.

(4) The interval of existence of this solution is the largest interval $I \subseteq \mathbb{R}$ *such that:*

(a) $t_0 \in I$, *and*

(b) $I \subseteq D$.

The general case. The general first-order linear case contains both the *direct integration case* and the *homogeneous case*. The trick with the integrating factor also works for the general case. We give an example first:

Example 3.3.7. Consider the first-order linear differential equation:

$$y' + \sin(t)y \;=\; \sin^3 t \tag{3.8}$$

The first thing to do is to compute the integrating factor:

$$\mu(t) \;=\; \exp\left(\int \sin t\,dt\right) \;=\; \exp(-\cos t)$$

Next we multiply both sides of (3.8) by $\mu(t)$ to obtain:

$$\mu(t)\big(y' + \sin(t)y\big) \;=\; \big(\exp(-\cos t)y\big)' \;=\; \sin^3 t \exp(-\cos t)$$

Integrating both sides yields:

$$\exp(-\cos t)y(t) \;=\; \int \sin^3 t \exp(-\cos t)\,dt \;=\; -4\exp(-\cos t)\cos^4(t/2) + C$$

Solving for $y(t)$ gives us the general solution:

$$y(t) \;=\; -4\cos^4(t/2) + C\exp\cos t$$

The general case works much the same way:

Theorem 3.3.8. *Suppose* $f : D \to \mathbb{R}$ *and* $g : E \to \mathbb{R}$ *are continuous functions with nice domains* $D, E \subseteq \mathbb{R}$ *and consider the differential equation*

(i) $y' + f(t)y = g(t)$

(1) Define the ***integrating factor*** *to be the function* $\mu : D \to \mathbb{R}$ *given by:*

$$\mu(t) \;:=\; \exp\left(\int f(t)\,dt\right)$$

(here $\int f(t)\,dt$ *can be any antiderivative of* $f(t)$*, the constant of integration does not matter). Then we can multiply (i) by* μ *to obtain:*

$$\mu(t)\big(y' + f(t)y\big) \;=\; \big(\mu(t)y\big)' \;=\; \mu(t)g(t).$$

(2) Then general solution of (i) is then given by:

$$y(t) \;=\; y(t;C) \;=\; \frac{1}{\mu(t)}\int \mu(t)g(t)\,dt + \frac{C}{\mu(t)}$$

Furthermore, suppose we are also given an initial condition

(ii) $y(t_0) = y_0$, *where* $t_0 \in D \cap E$ *and* $y_0 \in \mathbb{R}$.

(3) Then the initial value problem (i)+(ii) has the unique solution:

$$y(t) \;=\; \frac{1}{\mu(t)} \int_{t_0}^{t} \mu(s) g(s)\, ds + \frac{y_0}{\mu(t)}$$

where $\mu(t) := \exp(\int_{t_0}^{t} f(s)\, ds)$.

(4) The interval of existence of this solution is the largest interval $I \subseteq \mathbb{R}$ *such that:*

(a) $t_0 \in I$,

(b) $I \subseteq D$, *and*

(c) $I \subseteq E$.

PROOF. (1) First we will justify the key property of the integrating factor:

$$\mu(t)\big(y' + f(t)y\big) \;=\; \big(\mu(t)y\big)'$$

Note that:

$$\begin{aligned}
\big(\mu(t)y\big)' &= \mu(t)y' + \mu'(t)y \quad \text{by the product rule 2.3.4(2)} \\
&= \mu(t)y' + \frac{d}{dt}\left[\exp\left(\int f(t)\, dt\right)\right] y \\
&= \mu(t)y' + \exp\left(\int f(t)\, dt\right) \frac{d}{dt}\left[\int f(t)\, dt\right] y \quad \text{by the Chain Rule 2.3.6} \\
&= \mu(t)y' + \mu(t)f(t)y \quad \text{by Theorem 2.4.10} \\
&= \mu(t)\big(y' + f(t)y\big)
\end{aligned}$$

(2 part 1) Next, we will check that for every $C \in \mathbb{R}$, the function $y(t;C)$ is a solution. Since $\mu(t)$ is a function which is everywhere nonzero, it follows that $y(t;C)$ is a solution of

$$y' + f(t)y \;=\; g(t)$$

if and only if $y(t;C)$ is a solution of

$$(\dagger) \qquad \mu(t)\big(y' + f(t)y\big) \;=\; \mu(t)g(t).$$

We will verify that $y(t;C)$ is indeed a solution of (†). Note that:

$$\begin{aligned}
\mu(t)\big(y'(t;C) + f(t)y(t;C)\big) &= \big(\mu(t)y(t;C)\big)' \quad \text{by (1)} \\
&= \left(\int \mu(t)g(t) + C\right)' \\
&= \mu(t)g(t) \quad \text{by Theorem 2.4.10}
\end{aligned}$$

This verifies part (1) of Definition 3.2.3. We will return to verifying part (2) of the definition later.

(3 part 1) We now verify that

$$y(t) \;=\; \frac{1}{\mu(t)} \int_{t_0}^{t} \mu(s) g(s)\, ds + \frac{y_0}{\mu(t)}$$

is a solution to the initial value problem (i)+(ii). It is clear that $y(t)$ is a solution to (i) since it is a particular instance of the general solution in (2). To verify (ii),

we notice first that:

$$\begin{aligned}\mu(t_0) &= \exp\left(\int_{t_0}^{t_0} f(s)\,ds\right)\\ &= \exp(0) \quad \text{by Definition 2.4.1}\\ &= 1.\end{aligned}$$

Next, we observe:

$$\begin{aligned}y(t_0) &= \frac{1}{\mu(t_0)}\int_{t_0}^{t_0}\mu(s)g(s)\,ds + \frac{y_0}{\mu(t_0)}\\ &= \int_{t_0}^{t_0}\mu(s)g(s)\,ds + y_0\\ &= 0 + y_0 \quad \text{by Definition 2.4.1}\\ &= y_0.\end{aligned}$$

Thus $y(t)$ is a solution to the initial value problem (i)+(ii). We will prove uniqueness below.

(4) First observe that the interval $I \subseteq D$ is the largest possible interval which contains t_0 which we could hope to have as the domain of the solution. This is because the differential equation (i) is only defined on the set $D \cap E$ (the on which both coefficient functions f and g are defined).

(2 part 2) and (3 part 2) are taken care of by the next lemma. □

Lemma 3.3.9. *Suppose $f : D \to \mathbb{R}$ and $g : E \to \mathbb{R}$ are continuous functions with nice domains $D, E \subseteq \mathbb{R}$. Suppose that $y_0, y_1 : I \to \mathbb{R}$ are two differentiable functions such that:*

(a) $I \subseteq \mathbb{R}$ is an interval contained in both D and E,
(b) for $i = 0, 1$, $y_i'(t) + f(t)y_i(t) = g(t)$ for every $t \in I$, i.e., y_0 and y_1 are both solutions to the differential equation:

$$y' + f(t)y = g(t)$$

Then:

(1) there exists a constant $C \in \mathbb{R}$ such that for every $t \in I$,

$$y_0(t) = y_1(t) + \frac{C}{\mu(t)}$$

where $\mu(t) = \exp(\int f(t)\,dt)$.
(2) Furthermore, if there is $t_0 \in I$ such that $y_0(t_0) = y_1(t_0)$, then $C = 0$ and so for every $t \in I$, $y_0(t) = y_1(t)$.

Proof. It follows from (b) that for every $t \in I$,

$$(y_0 - y_1)'(t) + f(t)(y_0 - y_1)(t) = 0.$$

Multiplying both sides by $\mu(t)$ yields for every $t \in I$:

$$\mu(t)\big((y_0 - y_1)'(t) + f(t)(y_0 - y_1)(t)\big) = 0$$

which we can rewrite as:

$$\big(\mu(t)(y_0 - y_1)(t)\big)' = 0$$

for every $t \in I$. Since I is an interval, by Corollary 2.3.7 there is a constant $C \in \mathbb{R}$ such that for every $t \in I$:

$$\mu(t)(y_0 - y_1)(t) \;=\; C.$$

Thus for every $t \in I$,

$$y_0(t) \;=\; y_1(t) + \frac{C}{\mu(t)}.$$

This establishes (1). For (2), suppose there is $t_0 \in I$ such that $y_0(t_0) = y_1(t_0)$. Plugging in t_0 into the above equation then yields:

$$y_0(t_0) \;=\; y_1(t_0) + \frac{C}{\mu(t_0)}$$

which simplifies to

$$0 \;=\; \frac{C}{\mu(t_0)}.$$

This gives us $C = 0$. In particular, for every $t \in I$, we have

$$y_0(t) \;=\; y_1(t).$$

This establishes (2). □

Remark about absolute values in the integrating factor. In this subsection we make a few remarks about the role of absolute values in the integrating factor $\mu(t)$ which appears when computing a solution of a first-order linear differential equation. We begin with a soft rule-of-thumb:

Rule of Thumb 3.3.10. *If there are absolute values which arise in*

$$\mu(t) \;=\; \exp\left(\int f(t)\,dt\right)$$

as a result of an expression $\ln|\cdots|$ *arising in* $\int f(t)\,dt$*, then these absolute values can be safely removed in the final expression for* $\mu(t)$*.*

TLDR EXPLANATION. Suppose we are looking at the first-order linear differential equation:

$$y' + f(t)y \;=\; g(t)$$

The only relevant property that we need an integrating factor $\mu(t)$ to satisfy is that it simplifies the lefthand side:

$$\mu(t)\big(y' + f(t)y\big) \;=\; \big(\mu(t)y\big)' \tag{$\dagger$}$$

However, if $\mu(t)$ satisfies (†), then so does $-\mu(t)$:

$$-\mu(t)\big(y' + f(t)y\big) \;=\; \big(-\mu(t)y\big)'$$

since this amounts to multiplying (†) through by -1. Now suppose that $\mu(t) = |u(t)|$ for some differentiable function $u(t)$. Then by definition,

$$\mu(t) \;=\; \begin{cases} u(t) & \text{if } u(t) > 0 \\ -u(t) & \text{if } -u(t) < 0 \end{cases}$$

The claim is that the function $u(t)$ (i.e., μ without the absolute values) can serve as an integrating factor. This is essentially because:

$$u(t) \;=\; \begin{cases} \mu(t) & \text{if } u(t) > 0 \\ -\mu(t) & \text{if } u(t) < 0 \end{cases}$$

Since both $\mu(t)$ and $-\mu(t)$ work perfectly well as integrating factors, it follows that in all cases, the function $u(t)$ works as an integrating factor. □

We hesitate to call 3.3.10 a "Fact" or "Theorem" because this would require a complete investigation into all possible ways that an absolute value could show up in a formula for an antiderivative of an elementary function. However, we will give a justification as to why dropping absolute value signs is allowed and what we are actually doing to the integrating factor when we do drop the absolute value signs. For this discussion, we first make more precise what we mean by an *integrating factor*:

Definition 3.3.11. Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$ and $I \subseteq D$ is a nice subset of D. We call a differentiable function $\mu : I \to \mathbb{R}$ an **integrating factor for $y' + fy$ on I** if:

(1) $\mu(t) \neq 0$ for every $t \in I$, and
(2) for every differentiable function $y : I \to \mathbb{R}$, the following equality holds:

$$\mu(t)\big(y'(t) + f(t)y(t)\big) \;=\; \big(\mu(t)y(t)\big)'$$

for every $t \in I$.

Certainly, the integrating factors we've been using:

$$\mu(t) \;:=\; \exp\left(\int f(t)\,dt\right)$$

satisfy the definition of an *integrating factor* according to Definition 3.3.11. But an integrating factor is by no means unique. Indeed, we are free to multiply an integrating factor by any nonzero constant and it remains a perfectly valid integrating factor:

Observation 3.3.12. *Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$, $I \subseteq D$ is a nice subset of D, and $\mu : I \to \mathbb{R}$ is an integrating factor for $y' + fy$ on I. Then for any nonzero constant $\alpha \in \mathbb{R}$ ($\alpha \neq 0$), the function $\alpha\mu : I \to \mathbb{R}$ is also an integrating factor for $y' + fy$ on I.*

However, we have a little bit more freedom in modifying our integrating factors than just multiplying everything through by nonzero constants. For instance, consider the differential equation:

$$y' + \frac{1}{t}y \;=\; 0$$

We find that an integrating factor is $\mu(t) = \exp(\int dt/t) = |t|$. However, 3.3.10 claims that we can switch to using $\tilde{\mu}(t) = t$ as an integrating factor. The modification from $\mu(t)$ to $\tilde{\mu}(t)$ is more involved than just scaling $\mu(t)$ by a nonzero constant. First, note that in this example, $f(t) = 1/t$ and so $f : (-\infty, 0) \cup (0, +\infty) \to \mathbb{R}$ does not have 0 in its domain, so we are also considering $\mu(t) = |t|$ also to be a function $\mu : (-\infty, 0) \cup (0, +\infty) \to \mathbb{R}$ without zero in its domain. Furthermore, note that:

$$\mu(t) \;=\; \begin{cases} t & \text{if } t > 0 \\ -t & \text{if } t < 0 \end{cases} \quad \text{and} \quad \tilde{\mu}(t) \;=\; \begin{cases} t & \text{if } t > 0 \\ t & \text{if } t < 0 \end{cases}$$

In other words, to change $\mu(t)$ into $\tilde{\mu}(t)$, we had to multiply $\mu(t)$ by -1 on the $(-\infty, 0)$ portion of its domain, and keep $\mu(t)$ the same on the $(0, +\infty)$ portion of its domain. The reason this type of "selective" multiplication of $\mu(t)$ is allowed is

because $(-\infty, 0)$ and $(0, +\infty)$ are not connected to each other, so we don't have to worry about the portion of $\tilde{\mu}$ on $(-\infty, 0)$ joining up nicely with the portion of $\tilde{\mu}$ on $(0, +\infty)$. This is an instance of the following general observation:

Observation 3.3.13. *Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$, and suppose $\mu : D \to \mathbb{R}$ is an integrating factor for $y' + fy$ on D. Furthermore:*

(1) Suppose the domain $D = I_1 \cup I_2 \cup I_3 \cup \cdots$ is a union of disconnected intervals I_k (i.e., there is no $i \neq j$ and $a < b \in \mathbb{R}$ such that $[a, b] \subseteq I_i \cup I_j$), and

(2) Suppose $\alpha_1, \alpha_2, \alpha_3, \ldots$ is a sequence of nonzero constants from $\mathbb{R}$.

Then the function $\tilde{\mu} : D \to \mathbb{R}$ defined by:

$$\tilde{\mu}(t) \;:=\; \alpha_k \mu(t) \quad \text{if } t \in I_k$$

is also an integrating factor for $y' + fy$ on D.

We now arrive at a more precise version of 3.3.10:

Observation 3.3.14. *Suppose $f : D \to \mathbb{R}$ is a continuous function with a nice domain $D \subseteq \mathbb{R}$, and suppose*

$$\mu(t) \;:=\; \exp\left(\int f(t)\,dt\right) \;=\; \big|u(t)\big| \quad \text{for every } t \in D$$

where $u : D \to \mathbb{R}$ is some differentiable function. Then:

(1) for every $t \in D$, $u(t) \neq 0$,

(2) the sets,

$$D_1 \;:=\; \big\{t \in D : u(t) > 0\big\} \quad \text{and} \quad D_2 \;:=\; \big\{t \in D : u(t) < 0\big\}$$

are disconnected and $D = D_1 \cup D_2$, and thus

(3) the function $\tilde{\mu} : D \to \mathbb{R}$ defined by

$$\tilde{\mu}(t) \;:=\; u(t)$$

for every $t \in D$ is also an integrating factor of $y' + fy$.

JUSTIFICATION. (1) is clear because $\mu(t)$ is defined as an exponential of a certain function, and exp never takes the value zero.

(2) Suppose towards a contradiction that there is an interval $[a, b] \subseteq D$ such that $a \in D_1$ and $b \in D_2$ (the other case is similar). Then since $u : [a, b] \to \mathbb{R}$ is differentiable, and hence continuous, by the Intermediate Value Theorem 2.2.6 there is $y \in (a, b)$ such that $u(y) = 0$. This contradicts (1). Thus D_1 and D_2 are disconnected. The claim that $D = D_1 \cup D_2$ also follows from (1).

(3) is an application of Observation 3.3.13. In order to obtain $\tilde{\mu}$ from μ, on every interval $I \subseteq D_1$, we can keep μ the same, and on every interval $J \subseteq D_2$, we can multiply μ by -1. □

Remark 3.3.15. In general, you only need to worry about absolute value signs (and whether to drop them) when computing the general solution of a first-order linear differential equation. For an initial value problem, you use the precise integrating factor:

$$\mu(t) \;:=\; \exp\left(\int_{t_0}^{t} f(s)\,ds\right)$$

where t_0, t are both included in the same interval in the domain of f. Since your attention is restricted to this interval, the context should tell you, when faced with $|u(t)|$, whether to treat this as $u(t)$ or $-u(t)$ (depending on whether $u(t_0) > 0$ or $u(t_0) < 0$); only one of them can happen on an interval in the domain of f which contains t_0.

We now give a very carefully worked out example, where we show how to apply the above discussion on absolute values. In general, when you are doing computations, you are free to drop absolute values in this context without justification *provided that you still get the full correct answer.*

Example 3.3.16. Consider the following initial value problem:

(1) $y' + \tan(t)y = \sec(t)$
(2) $y(0) = 5$.

Find the general solution to (i) and the particular solution to (i)+(ii).

SOLUTION. First notice that the domain of $f(t) = \tan(t)$ and $g(t) = \sec(t)$ is

$$D \;:=\; \text{domain}(\tan t) \;=\; \text{domain}(\sec t) \;=\; \bigcup_{k\in\mathbb{Z}} \Big(\frac{\pi}{2} + \pi k, \frac{\pi}{2} + \pi(k+1)\Big)$$

i.e., the domain is all of $\mathbb{R}$ except points of the form $\pi/2 + \pi k$, where $k \in \mathbb{Z}$. Next we compute the usual integrating factor:

$$\mu(t) \;:=\; \exp\left(\int \tan t\, dt\right) \;=\; \exp \ln|\sec t| \;=\; |\sec t|.$$

The domain of $\mu(t)$ is the same as the domain of $\tan t$ and $\sec t$ above ($= D$). Furthermore, note that

$$D_1 \;:=\; \{t \in D : \sec t > 0\} \;=\; \bigcup_{k\in\mathbb{Z},k \text{ odd}} \Big(\frac{\pi}{2} + \pi k, \frac{\pi}{2} + \pi(k+1)\Big)$$

$$D_2 \;:=\; \{t \in D : \sec t < 0\} \;=\; \bigcup_{k\in\mathbb{Z},k \text{ even}} \Big(\frac{\pi}{2} + \pi k, \frac{\pi}{2} + \pi(k+1)\Big)$$

As we see, the intervals in D_1 are not connected to the intervals in D_2. Thus we can define $\tilde{\mu} : D \to \mathbb{R}$ by

$$\tilde{\mu}(t) \;:=\; \begin{cases} \mu(t) & \text{if } t \in D_1 \\ -\mu(t) & \text{if } t \in D_2 \end{cases} \;=\; \sec t$$

for every $t \in D$. By Observation 3.3.13, we know that $\tilde{\mu}(t) = \sec t$ also works as an integrating factor, so we will use that instead. Continuing on with the problem, we multiply (i) through by $\tilde{\mu}$ to obtain:

$$\big(\sec(t)y\big)' \;=\; \sec^2 t$$

Integrating both sides yields:

$$\sec(t)y \;=\; \tan t + C$$

where $C \in \mathbb{R}$ is an arbitrary constant. Thus the general solution[3] is:

$$y(t) \;=\; y(t;C) \;=\; \frac{\tan t + C}{\sec t}$$

on the domain D.

Next, we will solve the initial value problem (i)+(ii) from scratch. Since $t_0 = 0$, we see that the interval of existence of the solution will be $(-\pi/2, \pi/2)$, so we can restrict our attention to this interval. First we compute the integrating factor (where $t \in (-\pi/2, \pi/2)$):

$$\begin{aligned}
\mu(t) \;&:=\; \exp\left(\int_0^t \tan s\, ds\right) \\
&=\; \exp\left(\ln|\sec s|\Big|_0^t\right) \\
&=\; \exp\left(\ln \sec s\Big|_0^t\right) \quad (*) \\
&=\; \exp\big(\ln\sec t - \ln\sec 0\big) \\
&=\; \exp\big(\ln\sec t - \ln 1\big) \\
&=\; \exp\big(\ln\sec t\big) \\
&=\; \sec t
\end{aligned}$$

where in step $(*)$ we removed the absolute value signs because $\sec s$ is positive at $s = 0$ (if the initial condition had $t_0 = \pi$ for instance, then we would have to replace $\ln|\sec s|$ with $\ln(-\sec s)$ in that step). Now that we have the integrating factor, we can proceed with the particular solution (which is only defined on the interval of existence $(-\pi/2, \pi/2)$):

$$\begin{aligned}
y(t) \;&=\; \frac{1}{\sec t}\int_0^t \sec^2 s\, ds + \frac{5}{\sec t} \quad \text{because } y_0 = 5 \\
&=\; \frac{\tan t}{\sec t} + \frac{5}{\sec t} \\
&=\; \frac{\tan t + 5}{\sec t}.
\end{aligned}$$

□

Mixing problems. We now discuss a practical application of first-order linear differential equations, the so-called *mixing problems*. We will introduce mixing problems with an example from [**1**] and an example from [**2**]. All mixing problems basically follow the same general outline, although the differential equations which show up might vary.

Example 3.3.17 (Constant volume example)**.** Suppose a tank contains 10L of brine solution (salt dissolved in water). Assume the initial concentration of salt is 100g/L. Another brine solution flows into the tank at a rate of 3L/min with a concentration of 400g/L. Suppose the mixture is well stirred and flows out of the tank at a rate of 3L/min. Let $y(t)$ denote the amount of salt in the tank at time t. Find $y(t)$.

[3]Technically speaking, the general solution would have a possible different constant $+C$ on each connected component $(\pi/2 + k\pi, \pi/e + (k+1)\pi)$ of the domain, however we are sweeping this point under the rug. See Remark 2.4.9.

SOLUTION. We are interested in solving for

$$y(t) \;=\; \text{amount of salt, units: g.}$$

We will determine the function $y(t)$ by setting up and solving a differential equation for $y'(t)$:

$$y'(t) \;=\; \text{rate of change in amount of salt, units: g/min}$$

The main equation we will use is the so-called **balance law**:

$$y'(t) \;=\; \text{rate in} - \text{rate out}$$

Note that $y'(t)$, the "rate in" and "rate out" all have units g/min, whereas the information given in the question has units of either g/L or L/min. Thus we will need to use the following dimensional analysis:

$$\frac{\text{amount of salt}}{\text{unit of time}} \;=\; \frac{\text{volume of brine}}{\text{unit of time}} \times \frac{\text{amount of salt}}{\text{volume of brine}}$$

We now will determine the "rate in" and "rate out":

Rate in: The brine flows in at a rate of 3L/min with a fixed concentration of 400g/L. Thus the rate in of salt is:

$$\text{rate in} \;=\; 3\text{L/min} \times 400\text{g/L} \;=\; 1200\text{g/min.}$$

Rate out: The brine flows out at a rate of 3L/min. The concentration of the brine in the tank changes, however, depending on the value of $y(t)$. Since the tank contains a constant volume of brine, the concentration at time t in the tank is

$$\text{concentration in tank} \;=\; \frac{y(t)}{10}\text{g/L}$$

and thus the rate out is:

$$\text{rate out} \;=\; 3\text{L/min} \times \frac{y(t)}{10}\text{g/L} \;=\; \frac{3y(t)}{10}\text{g/min}$$

IVP: We conclude that the differential equation that y satisfies is:

$$y'(t) \;=\; 1200 - \frac{3}{10}y(t)$$

which we recognize as a first-order linear differential equation:

$$y' + \frac{3}{10}y \;=\; 1200.$$

Furthermore, at time $t = 0$, we know that $y(0) = 100\text{g/L} \times 10\text{L} = 1000g$. To summarize, we need to solve the IVP:

(i) $y' + \frac{3}{10}y = 1200$,
(ii) $y(0) = 1000$.

Using the usual method, we find that the solution is:

$$y(t) \;=\; 4000 - 3000e^{-3t/10}$$

where the units of $y(t)$ is g (grams). □

Here is a similar example, except that in this example, the volume of solution in the tank changes, as a result of an imbalance between the rate in and rate out:

Example 3.3.18 (Nonconstant volume example)**.** Suppose a 600L tank is filled with 300L of pure water at time $t = 0$. A spigot is opened above the tank and a brine solution with concentration 1.5g/L begins flowing into the tank at a rate of 3L/min. Simultaneously, a drain is opened at the bottom of the tank allowing the solution to leave the tank at a rate of 1L/min. What will be the salt content in the tank at the precise moment that the volume of solution in the tank is equal to the tank's capacity (=600L)?

SOLUTION. We need to perform a similar analysis as in Example 3.3.17 to get the function $y(t)$, but we also need to know at what time t_{full} is the volume of solution in the tank equal to 600L. Let $V(t)$ be the volume in the tank (in units of L). Then the change in volume is also governed by a balance law:

$$V'(t) \;=\; \text{rate in} - \text{rate out} \;=\; 3\text{g/min} - 1\text{g/min} \;=\; 2\text{g/min}$$

and thus

$$V(t) \;=\; 2t + C.$$

Since $V(0) = 300$, we get that $C = 300$ and so $V(t) = 2t + 300$. This allows us to determine the time t_{full} at which the tank is full:

$$600 \;=\; V(t_{\text{full}}) \;=\; 2t_{\text{full}} + 300$$

and thus $t_{\text{full}} = 150$min (so the tank will be full at the 3-hour mark).

Next we determine the function $y(t)$, again using the balance law:

$$y'(t) \;=\; \text{rate in} - \text{rate out}$$

Rate in: We are given that the solution which flows in has a rate of 3L/min, and a constant concentration of 400g/L. Thus:

$$\text{rate in} \;=\; 1.5\frac{\text{g}}{\text{L}} \times 3\frac{\text{L}}{\text{min}} \;=\; 4.5\frac{\text{g}}{\text{min}}$$

So the rate in of salt is constant.

Rate out: We are given that the solution flows out at a constant rate of 1L/min. The concentration in the tank, however, depends on the amount of salt in the tank $y(t)$, as well as the volume of solution in the tank $V(t)$. Thus:

$$\begin{aligned}\text{rate out} \;&=\; \text{volume rate out} \times \text{concentration in tank} \\ &=\; 1\frac{\text{L}}{\text{min}} \times \frac{y(t)\text{g}}{V(t)\text{L}} \;=\; \frac{y(t)}{2t+300}\frac{g}{L}\end{aligned}$$

Thus our differential equation for $y(t)$ is:

$$y' \;=\; 4.5 - \frac{y}{2t+300}$$

and our initial value is $y(0) = 0$ (since the tank starts with pure water, with no salt). This is a first-order linear differential equation. The solution is:

$$y(t) \;=\; 450 + 3t - \frac{4500\sqrt{3}}{\sqrt{300+2t}}$$

And thus the salt content at $t_{\text{full}} = 150$ is:

$$y(150) \;=\; 450 + 3\cdot 150 - \frac{4500\sqrt{3}}{\sqrt{300+2\cdot 150}} \;\approx\; 582\text{g}. \qquad \square$$

Variation of parameters. In this subsection we summarize an alternative method of solving a first-order linear differential equation, the method of *variation of parameters*. From a raw computational standpoint, this method requires you to compute the same integrals you otherwise would compute using the usual method, and for this reason we will not spend much time on it. However, it illustrates a certain idea in solving differential equations which we will encounter again:

A solution to the homogeneous equation can be used

to find a solution to the inhomogeneous equation.

We will illustrate the method of *variation of parameters* first through an example, and then give some general statements. We will only look at finding the general solution, a particular solution to an IVP is found using the initial condition from the general solution in the usual way (solving for C).

Example 3.3.19. Find the general solution to the following differential equation:

$$y' + y = \exp(t) \tag{3.9}$$

SOLUTION. We will solve this using variation of parameters in multiple steps:

Step 1: *Get the general solution to the homogeneous equation:*

$$y' + y = 0.$$

For this we do the same thing as before, first compute the integrating factor:

$$\mu(t) = \exp\left(\int dt\right) = \exp(t)$$

Now we multiply the differential equation through by $\mu(t)$ to obtain:

$$\big(\exp(t)y\big)' = 0$$

and then integrate to get the homogeneous solution (which we call y_n):

$$\exp(t)y_h(t) = C$$

and thus the general solution is:

$$y_h(t) = C\exp(-t)$$

Step 2: *Replace C with an unknown function, plug this into (i), and solve for the unknown function.*

Essentially, we will guess that the solution to $y(t) = v(t)\exp(-t)$, where $v(t)$ is an unknown function we need to find. Since $\mu(t) = \exp(-t)$ is everywhere nonzero, every solution of (i) technically can be written in the form $v(t)\exp(-t)$ (i.e., if $y(t)$ is a solution of (i), then $v(t) := y(t)\exp(t)$ works). If $y(t) = v(t)\exp(-t)$, then $y'(t) = v'(t)\exp(-t) - v(t)\exp(-t)$. Plugging these things into (i) yields:

$$\begin{aligned} y' + y &= \exp(t) \\ v'(t)\exp(-t) - v(t)\exp(-t) + v(t)\exp(-t) &= \exp(t) \\ v'(t)\exp(-t) &= \exp(t) \\ v'(t) &= \exp(2t). \end{aligned}$$

Solving for $v(t)$ (by integrating), we get that

$$v(t) = \frac{1}{2}\exp(2t) + C.$$

Thus the general solution to (i) is

$$y(t) \;=\; \left(\frac{\exp(2t)}{2} + C\right)\exp(-t) \;=\; \frac{\exp(t)}{2} + C\exp(-t). \qquad \square$$

Here are the steps in general for the method of *variation of parameters*:

Variation of Parameters 3.3.20. Consider the first-order linear differential equation:

$$y' + f(t) \;=\; g(t) \tag{3.10}$$

with corresponding homogeneous equation:

$$y' + f(t) \;=\; 0. \tag{3.11}$$

The method of **variation of parameters** to solve 3.10 consists of:

(1) First find the solution $y_h(t)$ to the homogeneous equation 3.11:
$$y_h(t) \;=\; \exp\left(-\int f(t)\,dt\right) \;=\; \frac{1}{\mu(t)}$$
where $\mu(t) = \exp(\int f(t)\,dt)$ is the usual integrating constant.

(2) Either substitute $y = v(t)y_h(t)$ into 3.10 and solve for $v(t)$, or else directly solve:
$$v' \;=\; \frac{g(t)}{y_h(t)}$$
with direct integration. The general solution will contain a constant of integration C.

(3) Write down the general solution to 3.11:
$$y(t) \;=\; v(t)y_h(t).$$

Note that in steps (1) and (2) you are basically performing the same two integrations that you do in the usual method of solving first-order linear differential equations. Thus not much is gained from choosing to use variation of parameters, except perhaps another point of view.

3.4. Implicit equations and differential forms

In this section we recall some facts from calculus about implicit equations and introduce the auxiliary tool of *differentials* and *differential forms*. By way of motivation, recall that we are ultimately interested in this chapter in solving explicit differential equations:

$$y' \;=\; F(t, y)$$

These equations can in general be much nastier than the first-order linear differential equations we studied in Section 3.3. The reason is because in general the two-variable function $F(t, y)$ might entangle the variables t and y together in some more complicated way than just "$-f(t)y + g(t)$". In calculus, we are used to most of the time y being an explicit function of t, i.e., $y = h(t)$ for some one-variable function h. However, this is ultimately a very special case and rather restrictive. Consequently:

We must abandon our desire for
y to always be an explicit function of t.

Instead, we will work with *implicitly defined equations*:

Definition 3.4.1. A **implicit equation** is a relation which can be written in the form:

$$F(t, y) \;=\; 0$$

where F is a function of two[4] variables. Given a two-variable function $F(t, y)$ and a constant $C \in \mathbb{R}$, we call the implicit equation:

$$F(t, y) \;=\; C$$

a **level set** of F.

Here is a very natural example of an implicit equation:

Example 3.4.2 (Circles)**.** Consider the function:

$$F(t, y) \;:=\; t^2 + y^2$$

Then for $C \in \mathbb{R}$, the level set:

$$t^2 + y^2 \;=\; C^2$$

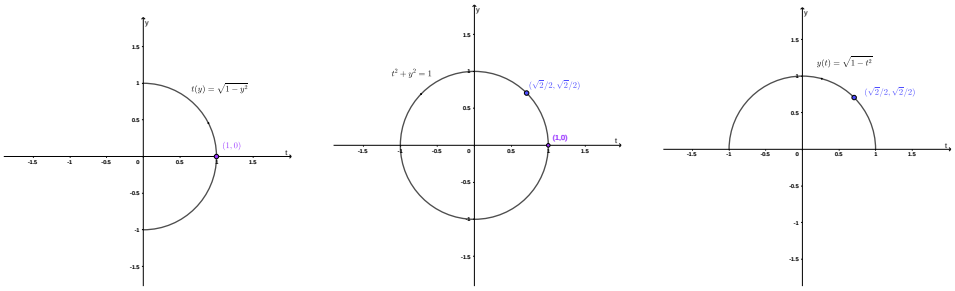
is the implicit equation which defines the circle of radius $|C|$ in the ty-plane. If $C \neq 0$, then the graph of $t^2 + y^2 = C^2$ is *not* a function since it fails the vertical line test. However, if we are interested in a certain point, say $(\sqrt{2}/2, \sqrt{2}/2)$ on the circle $t^2 + y^2 = 1$, then we can obtain an explicit function:

$$y(t) \;=\; \sqrt{1 - t^2}, \quad y : [-1, 1] \to \mathbb{R}$$

which passes through this point, and matches up with the top half of the full circle. If we are instead interested in the point $(1, 0)$, then we can instead look at the function:

$$t(y) \;=\; \sqrt{1 - y^2}, \quad t : [-1, 1] \to \mathbb{R}$$

which passes through this point and matches up with the right half of the full circle. We illustrate this example in Figure 3.3.



(A) Explicit equation (B) Implicit equation (C) Explicit equation

FIGURE 3.3. Implicit equation versus explicit equations for a circle

This illustrates in general how implicit equations work: implicit equations are *not* functions, but given a certain point (t_0, y_0) on the equation, there will be some function $y(t)$ or $t(y)$ which passes through the point and satisfies the equation.

[4]This definition generalizes to more than two variables, but we will restrict our attention to two variables in this section.

Question 3.4.3. *We know how to compute the derivative of an explicit function $y = f(t)$. The derivative is again an explicit function $dy/dt = f'(t)$. How do you "take the derivative" of an implicit equation $F(t, y) = 0$, and what type of object is "the derivative"?*

ANSWER. "The derivative" of an implicit equation $F(t, y) = 0$ is a brand new type of object, called a *differential form*:

Definition 3.4.4. A **differential form** is a formal expression of the form:

$$P(t, y)\, dt + Q(t, y)\, dy$$

where P, Q are two-variable functions and dt and dy are meaningless placeholders associated to the variables t and y called **differentials**. Differential forms can be added together in the natural way, and you can multiply them (from the left) by arbitrary functions $R(t, y)$.

The right notion of "taking the derivative" here is to compute the *differential* of $F(t, y)$:

Definition 3.4.5. Given a two-variable function $F(t, y)$, the **differential** of F (notation: dF) is the differential form:

$$dF \;:=\; \frac{\partial F}{\partial t}(t, y)\, dt + \frac{\partial F}{\partial y}(t, y)\, dy$$

Ultimately, we don't have to fully understand what the differential *really does* or what a differential form *really is*. We just need to know how to use them for certain types of computations. For us differential forms will appear as transient objects which make our calculations easier (for instance, see Example 3.4.6), especially when working with implicit equations and general first-order explicit differential equations. If you like, you can think of the differential dF as a "storage device" which contains all the "derivative information" associated with $F(t, y)$. □

We give an application of how you can use differential forms to compute implicit derivatives:

Example 3.4.6 (Implicit derivatives)**.** Consider the implicitly defined equation $t^2+y^2-1 = 0$ (circle of radius 1 in the ty-plane) and the point $(\sqrt{2}/2, \sqrt{2}/2)$. What is the derivative dy/dt of the implicitly defined function at the point $(\sqrt{2}/2, \sqrt{2}/2)$?

SOLUTION. One way to do this is to first notice that $(\sqrt{2}/2, \sqrt{2}/2)$ lies on the upper half-circle, so it is a point on the graph of the *explicit* function:

$$y(t) \;=\; \sqrt{1 - t^2}$$

Then we can compute:

$$\frac{dy}{dt}(t) \;=\; -\frac{t}{\sqrt{1 - t^2}}$$

and then plug in $t = \sqrt{2}/2$:

$$\frac{dy}{dt}\left(\frac{\sqrt{2}}{2}\right) \;=\; -\frac{\sqrt{2}/2}{\sqrt{1/2}} \;=\; -1.$$

This seems like an annoying way to answer this question because you have to:

(1) First, find an explicit function $y(t)$ which goes through the point and agrees with the implicit equation. This can sometimes be very hard or impossible to do exactly.
(2) Second, take the derivative of said explicit function. In our case, it also was annoying because we had to deal with the derivative of a square-root.

Here is a better way to do it:
First: Compute the differential of the equation $t^2 + y^2 - 1 = 0$. This will be:

$$2t\,dt + 2y\,dy \;=\; 0.$$

Second: "Solve" for dy/dt: since

$$2t\,dt + 2y\,dy \;=\; 0,$$

we can subtract $2t\,dt$ from both sides:

$$2y\,dy \;=\; -2t\,dt$$

and then divide both sides by $2y$ and "divide" both sides by dt:

$$\frac{dy}{dt} \;=\; -\frac{2t}{2y} \;=\; -\frac{t}{y}$$

Third: Plug in the point of interest:

$$\frac{dy}{dt} \;=\; -\frac{\sqrt{2}/2}{\sqrt{2}/2} \;=\; -1.$$

Although we write "solve" and "divide", we aren't actually doing anything sketchy. Given a correct and careful definition of *differential forms* and *differential* (which we won't go into), all of these steps are completely legitimate. Hopefully you are convinced that this is a much easier way to answer the question.

Another benefit of the differential form is that it is in some sense "coordinate neutral". For instance, suppose we asked a followup question: what is the derivative dt/dy at the point $(1, 0)$? Then we could just take the differential and "solve" for dt/dy in the same way:

$$\frac{dt}{dy} \;=\; -\frac{y}{t}$$

and so at $(1, 0)$:

$$\frac{dt}{dy} \;=\; -\frac{0}{1} \;=\; 0. \qquad \square$$

In some sense, the general process we will learn for solving differential equation $y' = F(t, y)$ is just this process in reverse (with a few more complications).

3.5. Separable and exact differential equations

In this section we will see how to essentially do the process in Example 3.4.6 in reverse, in order to solve an explicit first-order differential equation. As we will see, a much larger family of differential equations (beyond just the first-order linear ones) can be solved with this method. However, this method doesn't always guarantee an exact solution because in the worst case it requires you to solve a *partial* differential equation (PDE) which can be hard or impossible to solve exactly.

Obtaining a differential form equation. Given an explicit first-order linear differential equation

$$\frac{dy}{dt} = f(t,y) \tag{3.12}$$

the first step is to rewrite this as a differential form equation:

$$P(t,y)\,dt + Q(t,y)\,dy = 0 \tag{3.13}$$

This can be done with the following steps:

(Step 1) "Multiply" both sides of (3.12) by dt to get:

$$dy = f(t,y)\,dt$$

(Step 2) Subtract from both sides $f(t,y)\,dt$ to get:

$$-f(t,y)\,dt + dy = 0$$

(Step 3) If necessary, multiply both sides by some carefully chosen *integrating factor* $\mu(t,y)$:

$$-f(t,y)\mu(t,y)\,dt + \mu(t,y)\,dy = 0$$

Step 3 is the most important step, as this puts the differential form equation into a form we can "integrate" (i.e., compute an inverse of the differential d). We will see through examples some heuristics for how to do this for certain families of functions. We will also show how to check if the differential form equation can be solved. In the worst case, however, finding the right integrating factor $\mu(t,y)$ requires solving a PDE.

Separable differential equations. As a warmup, we will study a family of equations for which this process always works, the so-called *separable differential equations*:

Definition 3.5.1. A **separable equation** is an explicit first-order differential equation of the form:

(i) either

$$\frac{dy}{dt} = f(t)g(y)$$

(ii) or

$$\frac{dy}{dt} = \frac{f(t)}{g(y)}$$

where f, g are one-variable functions. Note that every equation of the form (ii) is also an equation of the form (i):

$$\frac{dy}{dt} = f(t)\left(\frac{1}{g(y)}\right) = f(t)h(t)$$

where $h = 1/g$. Thus we will restrict our attention to equations of the form (i).

The reason that a separable equation is called "separable", is because we can *separate the variables t and y* when performing Steps 1-3 above. Here are some examples:

Example 3.5.2. Here are some examples of separable equations and the corresponding "separated" differential form equation:

(1) $\frac{dy}{dt} = ty$. In this case, Step 1 and Step 2 yield:

$$-ty\,dt + dy \;=\; 0.$$

Now multiply both sides by $1/y$ to obtain:

$$-t\,dt + \frac{dy}{y} \;=\; 0.$$

(2) $\frac{dy}{dt} = e^{t-y}$. Recognize this equation as $dy/dt = e^t e^{-y}$. Then Step 1 and Step 2 yield:

$$-e^t e^{-y}\,dt + dy \;=\; 0$$

Multiplying both sides by e^y then gives us:

$$-e^t\,dt + e^y\,dy \;=\; 0$$

(3) $\frac{dy}{dt} = ty + y$. Rewrite this as $dy/dt = (t+1)y$. Then get:

$$-(t+1)y\,dt + dy \;=\; 0$$

and multiplying by $1/y$ yields:

$$-(t+1)\,dt + \frac{dy}{y} \;=\; 0.$$

These examples show that in general a separable equation

$$\frac{dy}{dt} \;=\; f(t)g(y)$$

gives rise to the differential form equation

$$-f(t)\,dt + \frac{dy}{g(y)} \;=\; 0.$$

Since each differential dt and dy has as coefficient functions a one-variable function *in the same variable*, we can "integrate" this differential form equation using the following:

Observation 3.5.3. *Given a separated differential form equation:*

$$P(t)\,dt + Q(y)\,dy \;=\; 0 \tag{3.14}$$

Define the two-variable function:

$$F(t,y) \;:=\; \int P(t)\,dt + \int Q(y)\,dy$$

Then

$$dF \;=\; \frac{\partial}{\partial t}\left(\int P(t)\,dt\right)\,dt + \frac{\partial}{\partial y}\left(\int Q(y)\,dy\right)\,dy \;=\; P(t)\,dt + Q(y)\,dy$$

Thus, the implicit equation

$$F(t,y) \;=\; C$$

where $C \in \mathbb{R}$ is arbitrary, is the "general solution" to the differential form equation (3.14).

In other words, to integrate a separated differential form equation, you just compute two one-variable integrals, a dt-integral and a dy-integral. Each one gives you a constant of integration, but these constants of integration can be combined into one and put on the righthand side of the equation. We illustrate this with a few examples:

Example 3.5.4. Continuing with our examples from 3.5.2:

(1) Given our differential form equation

$$-t\,dt + \frac{dy}{y} \;=\; 0$$

we integrate both parts of the lefthand side separately to get:

$$\int -t\,dt + \int \frac{dy}{y} \;=\; -\frac{t^2}{2} + \ln|y| \;=\; C.$$

Thus the general solution, as an implicit equation, is:

$$-\frac{t^2}{2} + \ln|y| \;=\; C.$$

(2) Given our differential form equation

$$-e^t\,dt + e^y\,dy \;=\; 0$$

we integrate to get:

$$\int -e^t\,dt + \int e^y\,dy \;=\; -e^t + e^y \;=\; C.$$

Thus the general solution, as an implicit equation, is:

$$-e^t + e^y \;=\; C.$$

(3) Given our differential form equation

$$-(t+1)\,dt + \frac{dy}{y} \;=\; 0$$

we integrate to get

$$\int -(t+1)\,dt + \int \frac{dy}{y} \;=\; -\frac{(t+1)^2}{2} + \ln|y| \;=\; C$$

Thus the general solution, as an implicit equation, is:

$$-\frac{(t+1)^2}{2} + \ln|y| \;=\; C.$$

Here is a convention for this class involving separable (and also exact) equations below:

Convention 3.5.5. If we ask for the general solution to a separable or exact differential equation, you may leave the general solution in implicit form *unless* we specifically ask you to put it in explicit form, in which case you have to solve for y in terms of C. If a term $|y| = u(t;C)$ shows up, then this simplifies to $y = \pm u(t;C)$.

Example 3.5.6. We will continue with the three examples from Example 3.5.4, giving the general solution in explicit form:

(1) Our general solution in implicit form is $-t^2/2 + \ln|y| = C$. Solving for y yields:

$$\begin{aligned} \ln|y(t)| &= \frac{t^2}{2} + C \\ |y(t)| &= \exp\left(\frac{t^2}{2} + C\right) \\ y(t) &= \pm\exp\left(\frac{t^2}{2} + C\right) \quad \text{(general solution)} \end{aligned}$$

(2) Our general solution in implicit form is $-e^t + e^y = C$. Solving for y yields:

$$\begin{aligned} -e^t + e^y &= C \\ e^y &= e^t + C \\ y(t) &= \ln\left(e^t + C\right) \quad \text{(general solution)} \end{aligned}$$

Note: we do not put absolute values in the last step. The second equation $e^y = e^t + C$ tells us that $e^t + C$ *must* be positive. This places additional conditions on the constant C and the domain of the general solution (which will be a function of C) — something we will not bother with.

(3) Our general solution in implicit form is $-(t+1)^2/2 + \ln|y| = C$. Solving for y yields:

$$\begin{aligned} \ln|y| &= \frac{(t+1)^2}{2} + C \\ |y(t)| &= \exp\left(\frac{(t+1)^2}{2} + C\right) \\ y(t) &= \pm\exp\left(\frac{(t+1)^2}{2} + C\right) \quad \text{(general solution)} \end{aligned}$$

Here is the convention for initial value problems:

Convention 3.5.7. Suppose our separable or exact differential equation as implicit general solution:

$$F(t, y) = C$$

and we also have an initial condition $y(t_0) = y_0$. Then:

(1) First solve for C by noticing $C = F(t_0, y_0)$. If we do not explicitly ask for the particular solution in explicit form, then you may stop here.
(2) If we do ask for the explicit solution, then solve $F(t, y) = C$ (with the new exact value for C) for y, using the initial condition $y(t_0) = y_0$ anytime you have to make a choice (e.g., dealing with absolute values, or square roots). The interval of existence will be the largest interval which contains t_0 for which $y(t)$ is naturally defined.

Example 3.5.8. Find the particular solution (in explicit form) for the following initial value problem:

(i) $y' = ty$
(ii) $y(1) = 1$

SOLUTION. We have found the implicit general solution in Example 3.5.4 to be:

$$-\frac{t^2}{2} + \ln|y| \;=\; C$$

Solving for C yields:

$$C \;=\; -\frac{1^2}{2} + \ln|1| \;=\; -\frac{1}{2}.$$

Next we solve for $y(t)$:

$$\begin{aligned} -\frac{t^2}{2} + \ln|y(t)| &= -\frac{1}{2} \\ \ln|y(t)| &= \frac{t^2}{2} - \frac{1}{2} \\ |y(t)| &= \exp\left(\frac{t^2}{2} - \frac{1}{2}\right) \\ y(t) &= \exp\left(\frac{t^2}{2} - \frac{1}{2}\right) \end{aligned}$$

Here we needed to take the righthand side to be positive when we removed the absolute values because $y_0 = 1$ is positive. The interval of existence is all of $\mathbb{R}$ as this is the natural domain of the righthand function of t. □

We end our separable discussion with a remark about dividing by zero:

Remark 3.5.9. Suppose we have an initial value problem:

(i) $y' = f(t)g(y)$
(ii) $y(t_0) = y_0$.

and $g(y_0) = 0$. Then the constant function $y(t) = y_0$ for all t is a solution and the interval of existence is the largest possible interval which contains t_0 for which $f(t)$ is defined.

Exact differential equations. Now we move on to the general case:

$$y' \;=\; f(t,y) \tag{3.15}$$

where f is a two-variable function which might not be separable (i.e., it might not be of the form $f(t,y) = g(t)h(y)$). Recall that the first order of business is to translate equation (3.15) into a suitable differential form equation:

(Step 1) Rewrite (3.15) as $\frac{dy}{dt} = f(t,y)$.
(Step 2) "Multiply" both sides by dt, then add $-f(t,y)\,dt$ to both sides to obtain:

$$-f(t,y)\,dt + dy \;=\; 0$$

(Step 3) Multiply both sides by a carefully chosen *integrating factor* $\mu(t,y)$:

$$-f(t,y)\mu(t,y)\,dt + \mu(t,y)\,dy \;=\; 0.$$

This will give us a differential form equation:

$$P(t,y)\,dt + Q(t,y)\,dy \;=\; 0. \tag{3.16}$$

Of course, we have said nothing yet about how to find the integrating factor $\mu(t,y)$, or what it needs to do. Ultimately, to solve a differential form equation of the form (3.16), we need to find a so-called *potential function*:

Definition 3.5.10. A **potential function** for (3.16) is a two-variable function $F(t, y)$ such that

$$dF \;=\; \frac{\partial F}{\partial t}\,dt + \frac{\partial F}{\partial y}\,dy \;=\; P(t, y)\,dt + Q(t, y)\,dy,$$

i.e.,

(1) $\frac{\partial F}{\partial t} = P(t, y)$, and
(2) $\frac{\partial F}{\partial y} = Q(t, y)$.

In other words, a potential function is like an antiderivative of a differential form.

Unfortunately, not every differential form has a potential function. This begs the question:

Question 3.5.11. *When does the differential form $P(t, y)\,dt + Q(t, y)\,dy$ have a potential function?*

Answer. First, we will define what it means for a differential form to have a potential function:

Definition 3.5.12. Suppose $P, Q : D \to \mathbb{R}$ are continuous two-variable functions on a nice domain $D \subseteq \mathbb{R}^2$. We say that the differential form

$$P\,dt + Q\,dy$$

is **exact** if there exists a continuously differentiable function $F : D \to \mathbb{R}$ such that

$$dF \;=\; P\,dt + Q\,dy.$$

Next, we isolate a necessary condition (which is easily checkable) for a differential form to be exact. We will further assume that P and Q are *continuously differentiable* (this will be the case for all the functions we shall encounter). Suppose $F(t, y)$ is a potential function of $P(t, y)\,dt + Q(t, y)\,dy$, so F will have to have continuous second-order partial derivatives (in order for P and Q to have continuous first-order partial derivatives). Then by the Clairaut-Schwarz Theorem 2.3.13 it follows that:

$$\frac{\partial^2 F}{\partial t \partial y} \;=\; \frac{\partial^2 F}{\partial y \partial t}$$

Thus, since $\frac{\partial F}{\partial t} = P$ and $\frac{\partial F}{\partial y} = Q$, then this says that

$$\frac{\partial Q}{\partial t} \;=\; \frac{\partial P}{\partial y}$$

i.e., the partial derivatives of P and Q with respect to the *other* variable must be the same. This motivates the following definition:

Definition 3.5.13. Suppose $P, Q : D \to \mathbb{R}$ are continuously differentiable two-variable functions on a nice domain $D \subseteq \mathbb{R}^2$. We say that the differential form

$$P\,dt + Q\,dy$$

is **closed** if

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial y} \;=\; 0$$

i.e., if the lefthand side is the constant zero function.

Clearly, in order for a differential form to be exact, it must also be closed (which is a very easy condition to check). What about the converse? As it turns out, if we impose a natural condition on the domain D, then these two are equivalent:

Theorem 3.5.14. *Suppose $P, Q : I \times J \to \mathbb{R}$ are continuously differentiable functions and $I, J \subseteq \mathbb{R}$ are intervals (so the common domain of P and Q is a rectangle). Then the following are equivalent:*

(1) the differential form $P\,dt + Q\,dy$ is exact.
(2) the differential form $P\,dt + Q\,dy$ is closed.

This provides an answer to the original question, namely, if the functions P and Q are nice (continuously differentiable, which they always will be for us), and the domain is a rectangle, then the differential form $P\,dt + Q\,dy$ has a potential function iff $P\,dt + Q\,dy$ is closed, i.e., iff $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial t}$. □

Here is an example which shows how Theorem **??** can fail if the domain of P, Q is not a rectangle:

Example 3.5.15. Consider the differential form equation:

$$\frac{-y}{t^2 + y^2}\,dt + \frac{t}{t^2 + y^2}\,dy \;=\; 0.$$

Here the domain of the coefficient functions is $\mathbb{R}^2 \setminus \big\{(0,0)\big\}$, i.e., the entire ty-plane except the origin. It is easy to see that this differential form is closed:

$$\begin{aligned}
\frac{\partial}{\partial y}\left(\frac{-y}{t^2 + y^2}\right) &= \frac{y^2 - t^2}{(t^2 + y^2)^2} \\
\frac{\partial}{\partial t}\left(\frac{t}{t^2 + y^2}\right) &= \frac{y^2 - t^2}{(t^2 + y^2)^2}
\end{aligned}$$

However, the differential form is not exact. Assume towards a contradiction that there exists a potential function $F : \mathbb{R}^2 \setminus \big\{(0,0)\big\} \to \mathbb{R}$. Then on the one hand we would have

$$\int_0^{2\pi} \frac{d}{d\theta} F(\cos\theta, \sin\theta)\,d\theta \;=\; F(1,0) - F(1,0) \;=\; 0.$$

On the other hand, we have by the (multivariable) chain rule:

$$\begin{aligned}
\frac{d}{d\theta} F(\cos\theta, \sin\theta) &= \frac{\partial F}{\partial t} \cdot (-\sin\theta) + \frac{\partial F}{\partial y} \cdot \cos\theta \\
&= \frac{\sin\theta}{\cos^2\theta + \sin^2\theta} \cdot \sin\theta + \frac{\cos\theta}{\cos^2\theta + \sin^2\theta} \cdot \cos\theta \\
&= 1,
\end{aligned}$$

which implies that $\int_0^{2\pi} \frac{d}{d\theta} F(\cos\theta, \sin\theta)\,d\theta = 2\pi \neq 0$. This is a contradiction, and so no such potential function F can exist.

We now provide an example of checking whether a given differential form is closed (and also exact):

Example 3.5.16. (1) $(2t+y)\,dt + (t-6y)\,dy$. First we compute the partial derivatives $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial t}$:

$$\begin{aligned}\frac{\partial}{\partial y}(2t+y) &= 1\\ \frac{\partial}{\partial t}(t-6y) &= 1\end{aligned}$$

Thus $(2t+y)\,dt + (t-6y)\,dy$ is closed. Since both P and Q are defined on the rectangle $\mathbb{R}\times\mathbb{R}$, by Theorem 3.5.14 this differential form is exact, hence there exists a potential function for it.

(2) $(2t+\ln y)\,dt + ty\,dy$. First we compute the relevant partials:

$$\begin{aligned}\frac{\partial}{\partial y}(2t+\ln y) &= \frac{1}{y}\\ \frac{\partial}{\partial t}(ty) &= y\end{aligned}$$

Since these partial derivatives are not equal, the differential form is not closed, hence it is not exact.

The next order of business is to solve for a potential function of an exact differential form. This can be done with the following steps:

Finding a potential function of an exact differential form 3.5.17. Suppose the differential form $P(t,y)\,dt + Q(t,y)\,dy$ is exact. The solution to the differential form equation

$$P(t,y)\,dt + Q(t,y)\,dy = 0$$

is $F(t,y) = C$, where F is a potential function of $P(t,y)\,dt + Q(t,y)\,dy$. A potential function F can be found in the following steps:

(1) First solve $\frac{\partial F}{\partial t} = P$ by integrating with respect to t:

$$F(t,y) = \int P(t,y)\,dt + \phi(y) \tag{3.17}$$

where $\phi(y)$ is an unknown function of y only. Here $\phi(y)$ plays the role of "constant of integration", except that since we are considering partial derivatives and integrating with respect to t only, we have to allow our constant of integration to in fact be a function of y.

(2) Next, we need to find what $\phi(y)$ is. Since we know $\frac{\partial F}{\partial y} = Q(t,y)$, we can differential (3.17) with respect to y:

$$\frac{\partial}{\partial y}\int P(t,y)\,dt + \phi'(y) = Q(t,y),$$

and thus

$$\phi(y) = \int \left(Q(t,y) - \frac{\partial}{\partial y}\int P(t,y)\,dt\right) dy$$

(3) Now that we know what function $\phi(y)$ is, our general solution (in implicit form) is:

$$F(t,y) = C.$$

We give some examples as to how this process works:

Example 3.5.18. In each of the following examples, the differential form is exact and we will solve the indicated differential form equation.

(1) $(2t\sin y + y^3e^t)\,dt + (t^2\cos y + 3y^2e^t)\,dy = 0$. First we verify that the differential form is exact. Indeed:

$$\begin{aligned}\frac{\partial}{\partial y}(2t\sin y + y^3e^t) &= 2t\cos y + 3y^2e^t\\ \frac{\partial}{\partial t}(t^2\cos y + 3y^2e^t) &= 2t\cos y + 3y^2e^t\end{aligned}$$

Now we will find a potential function $F(t,y)$ for this differential form. First using that $\frac{\partial F}{\partial t} = 2t\sin y + y^3e^t$, we get that

$$F(t,y) = \int (2t\sin y + y^3e^t)\,dt + \phi(y) = t^2\sin y + y^3e^t + \phi(y)$$

for some unknown function $\phi(y)$ which is solely a function of y. Next we take the partial derivative of this F with respect to y and set it equal to $t^2\cos y + 3y^2e^t$:

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}(t^2\sin y + y^3e^t + \phi(y)) = t^2\cos y + 3y^2e^t + \phi'(y) = t^2\cos y + 3y^2e^t$$

Thus $\phi'(y) = 0$. Integrating with respect to y finally yields $\phi(y) = C$. Thus our potential function is:

$$F(t,y) = t^2\sin y + y^3e^t + C.$$

We conclude that our general solution is:

$$t^2\sin y + y^3e^t + C = 0.$$

Replacing C with $-C$, this general solution is equivalent to:

$$t^2\sin y + y^3e^t = C.$$

(2) $\big(1+(1+ty)e^{ty}\big)\,dt + (1+t^2e^{ty})\,dy = 0$. First we verify that the differential for is exact. Indeed:

$$\begin{aligned}\frac{\partial}{\partial y}\big(1+(1+ty)e^{ty}\big) &= (1+ty)e^{ty}t + te^{ty} = e^{ty}(2t+t^2y)\\ \frac{\partial}{\partial t}(1+t^2e^{ty}) &= 2te^{ty} + t^2e^{ty}y = e^{ty}(2t+t^2y)\end{aligned}$$

Now we will find a potential function for this differential form. First, using that $\frac{\partial F}{\partial t} = 1+(1+ty)e^{ty}$, we get that

$$F(t,y) = \int \big(1+(1+ty)e^{ty}\big)\,dt + \phi(y) = te^{ty} + t + \phi(y)$$

for some unknown function $\phi(y)$. Next, we take the partial derivative of this F with respect to y and set it equal to $1+t^2e^{ty}$:

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\big(te^{ty} + t + \phi(y)\big) = t^2e^{ty} + \phi'(y) = 1 + t^2e^{ty}$$

Thus $\phi'(y) = 1$. Integrating with respect to y yields $\phi(y) = y + C$. We conclude that our potential function is:

$$F(t,y) = te^{ty} + t + y + C$$

and thus our general solution is:

$$te^{ty} + t + y + C \;=\; 0.$$

The integrating factor $\mu(t, y)$. We have not said anything about the integrating factor yet. Its role is as follows:

The integrating factor makes a non-exact equation exact.

Specifically, here is the definition:

Definition 3.5.19. Suppose $P, Q : D \to \mathbb{R}$ are continuous on a nice domain $D \subseteq \mathbb{R}^2$. We say that a function $\mu : D \to \mathbb{R}$ is an **integrating factor** for the differential form equation

$$P(t, y)\, dt + Q(t, y)\, dy \;=\; 0$$

if

(i) $\mu(t, y) \neq 0$ for every $(t, y) \in D$, and
(ii) $\mu(t, y)P(t, y)\, dt + \mu(t, y)Q(t, y)\, dy$ is exact.

In particular, if $D \subseteq \mathbb{R}^2$ is a rectangle, then by Theorem 3.5.14 (ii) is satisfied if and only if:

$$\frac{\partial}{\partial y}\big(\mu(t, y)P(t, y)\big) \;=\; \frac{\partial}{\partial y}\big(\mu(t, y)Q(t, y)\big) \;=\; 0$$

i.e., if and only if:

$$\frac{\partial \mu}{\partial y}Q(t, y) + \mu(t, y)\frac{\partial Q}{\partial y} \;=\; \frac{\partial \mu}{\partial t}P(t, y) + \mu(t, y)\frac{\partial P}{\partial t} \tag{3.18}$$

In general, if a differential form equation is not-exact, then finding an integrating factor involves solving the partial differential equation (PDE) given in (3.18). This can be hard/impossible to do. For this reason, we will not study techniques for finding this integrating factor in this class. Here are the conventions for this class as to what you're expected to know how to do with regards to this integrating factor:

Convention 3.5.20. You need to know how to do the following things for this class:

(1) Be able to check if a differential form equation is exact, and solve it if it is exact.
(2) Given a non-exact differential form equation, and supplied with a valid integrating factor, you need to be able to use the integrating factor to solve the equation.

Here is an example of solving a non-exact differential form equation after being supplied with a valid integrating factor.

Example 3.5.21. Consider the differential form equation $(3t^2y+2ty+y^3)\, dt+(t^2+y^2)\, dy = 0$ and the integrating factor $\mu(t, y) = e^{3t}$. First note that the differential form is not exact:

$$\frac{\partial}{\partial y}(3t^2y + 2ty + y^3) \;=\; 3t^2 + 2t + 3y^2$$

$$\frac{\partial}{\partial t}(t^2 + y^2) \;=\; 2t$$

however, multiplying through by $\mu(t,y) = e^{3t}$ yields the differential form equation:

$$(3t^2y + 2ty + y^3)e^{3t}\,dt + (t^2 + y^2)e^{3t}\,dy \;=\; 0$$

which is exact:

$$\frac{\partial}{\partial y}(3t^2y + 2ty + y^3)e^{3t} \;=\; (3t^2 + 2t + 3y^2)e^{3t}$$

$$\frac{\partial}{\partial t}(t^2 + y^2)e^{3t} \;=\; (t^2 + y^2)3e^{3t} + 2te^{3t} \;=\; (3t^2 + 3y^2 + 2t)e^{3t}$$

Now we will solve for the potential function. Using $\frac{\partial F}{\partial t} = (3t^2y + 2ty + y^3)e^{3t}$ we get

$$F(t,y) \;=\; \int (3t^2y + 2ty + y^3)e^{3t}\,dt + \phi(y) \;=\; e^{3t}y(t^2 + \frac{y^2}{3}) + \phi(y)$$

Next, taking a partial derivative with respect to y and setting this equal to $(t^2 + y^2)e^{3t}$ yields:

$$\frac{\partial F}{\partial y} \;=\; e^{3t}(t^2 + y^2) + \phi'(y) \;=\; (t^2 + y^2)e^{3t}.$$

We conclude that $\phi'(y) = 0$. Integrating this with respect to y yields $\phi(y) = C$. We conclude that our potential function is:

$$F(t,y) \;=\; (t^2 + y^2)e^{3t} + C$$

and thus our general solution is:

$$(t^2 + y^2)e^{3t} + C \;=\; 0.$$

3.6. Existence and uniqueness theorems

We have already seen the full existence and uniqueness theorem for first-order linear differential equations (Theorem 3.3.8). In this section we will give statements of other existence and uniqueness theorems.

We have already given the relevant *existence and uniqueness theorem* for first-order linear differential equations in Theorem 3.3.8. The following is the corresponding statement for separable differential equations. Note that in general for separable differential equations, we are only guaranteed *local uniqueness*, i.e., a unique solution on a tiny interval I' which contains t_0 (provided $g(y_0) \neq 0$). At this level of generality, we can't really say what the largest possible interval of existence will be (unlike the statement of Theorem 3.3.8), although in practice you may be able to determine this when solving for the explicit solution to an IVP.

Existence and Uniqueness Theorem 3.6.1 (Separable case)**.** *Suppose $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ are continuous functions defined on intervals I and J. Consider the initial value problem:*

(i) $y' = f(t)g(y)$
(ii) $y(t_0) = y_0$*, where* $t_0 \in I$*,* $y_0 \in J$*.*

(1) If y_0 is not an endpoint of J and $g(y_0) \neq 0$, then
  (a) the initial value problem (i)+(ii) has a unique solution $y(t) : I' \to \mathbb{R}$, where $I' \subseteq I$ is some open interval containing t_0, and

(b) the solution to the initial value problem can be obtained by solving for y in the following equation:

$$\int_{y_0}^{y} \frac{ds}{g(s)} \;=\; \int_{t_0}^{t} f(s)\,ds.$$

(2) If $g(y_0) = 0$, then the constant function $y(t) = y_0$, $y : I \to \mathbb{R}$, is a solution to (i)+(ii), but it may not be unique.

We now present the main existence theorem for explicit first-order differential equations:

Existence Theorem 3.6.2 (General case)**.** *Suppose $f : I \times J \to \mathbb{R}$ is a continuous two-variable function defined on a rectangle $I \times J$ in the ty-plane (so $I, J \subseteq \mathbb{R}$ are intervals). Then given any point $(t_0, y_0) \in I \times J$, the initial value problem*

(i) $y' = f(t, y)$
(ii) $y(t_0) = y_0$

has a solution $y(t)$ define on some interval $I' \subseteq I$ which contains t_0. Furthermore, the solution will be defined at least until the solution curve $t \mapsto \big(t, y(t)\big)$ leaves the rectangle $I \times J$.

The following example illustrates what we mean by "leaving the rectangle":

Example 3.6.3. Consider the IVP:

(i) $y' = 1 + y^2$
(ii) $y(0) = 0$

For this differential equation, the function $f(t, y)$ is $f(t, y) = 1+y^2$, which is defined everywhere on the ty-plane. Thus we can consider its domain to be the rectangle $\mathbb{R} \times \mathbb{R}$. Solving this as a separable equation yields the solution $y(t) = \tan t$. The interval of existence is $(-\pi/2, \pi/2)$, since this is the interval in the domain of $\tan t$ which contains $t_0 = 0$. This agrees with the Existence Theorem 3.6.2 since $y(t)$ "leaves the rectangle" at $\pm\pi/2$ in the sense that it has vertical asymptotes at these t-values, so it shoots down/up to $\pm\infty$ at these points and "leaves" the ty-plane.

We also have the main uniqueness theorem for explicit first-order differential equations. Note that the uniqueness theorem requires stronger hypotheses than the existence theorem, so it holds in fewer situations.

Uniqueness Theorem 3.6.4 (General case)**.** *Suppose $f : I \times J \to \mathbb{R}$ is a continuous two-variable function defined on a rectangle $I \times J$ in the ty-plane (so $I, J \subseteq \mathbb{R}$ are intervals). Furthermore, suppose the partial derivative $\frac{\partial f}{\partial y}$ exists and is continuous on all of $I \times J$. Let $(t_0, y_0) \in I \times J$, and suppose we have have two solutions $y(t), \tilde{y}(t)$ to the same IVP:*

(1) $y'(t) = f\big(t, y(t)\big)$ *and* $\tilde{y}'(t) = f\big(t, \tilde{y}(t)\big)$ *for every t, and*
(2) $y(t_0) = y_0$ *and* $\tilde{y}(t_0) = y_0$.

Then for every t such that $\big(t, y(t)\big)$ and $\big(t, \tilde{y}(t)\big)$ remain in the rectangle $I \times J$, we have

$$y(t) \;=\; \tilde{y}(t).$$

One of the practical benefits of the Uniqueness Theorem 3.6.4 is that, *provided the hypotheses of 3.6.4 are satisfied*, then

Different solution curves cannot cross.

Here is a (somewhat exaggerated and contrived) example of this principle:

Example 3.6.5. Consider the differential equation:

$$y' = (y-10)\sin(x+y)e^{e^{x-y}} \tag{3.19}$$

and suppose that $\tilde{y} : I \to \mathbb{R}$ is a solution to the equation (3.19) on an interval I which contains 0. Furthermore, assume that $\tilde{y}(0) = 0$. Then $\tilde{y}(t) \leq 10$ for all $t \in I$.

JUSTIFICATION. First note that $\bar{y} : I \to \mathbb{R}$ defined by $\bar{y}(t) := 10$ for all $t \in I$ is also a solution of 3.19. Then $\bar{y}(0) = 10$ whereas $\tilde{y}(0) = 0$. Thus by the Uniqueness Theorem 3.6.4, there can be no $t_0 \in I$ such that $\bar{y}(t_0) = \tilde{y}(t_0)$. Thus the two differentiable (hence continuous) functions $\bar{y}$ and $\tilde{y}$ never intersect. Finally, since $\tilde{y}(0) < \bar{y}(0)$, it follows that for all $t \in I$ that $\tilde{y}(t) < \bar{y}(t) = 10$ (since these functions cannot intersect). □

Notice that the inequality established in Example 3.6.5 would be hard to establish directly without the Uniqueness Theorem, since the equation (3.19) looks hard/impossible to solve exactly.

3.7. Autonomous equations

In this final section, we will take a look at qualitative properties of solutions of the so-called *autonomous equations*:

Definition 3.7.1. A first-order differential equation is called an **autonomous equation** if it can be written in the form:

$$y' = f(y)$$

i.e., if the equation does not depend on the independent variable t.

Autonomous equations are a special case of separable equations, and hence could be solved using the methods from Section 3.5. However, we will be more interested in studying the *qualitative* properties of its solutions, i.e., saying as much as we can about the solutions without explicitly solving for them.

Example 3.7.2. Consider the autonomous equation

$$y' = (y+1)(y^2-9)$$

Below we have a sketch of the direction field along with several solutions curves:

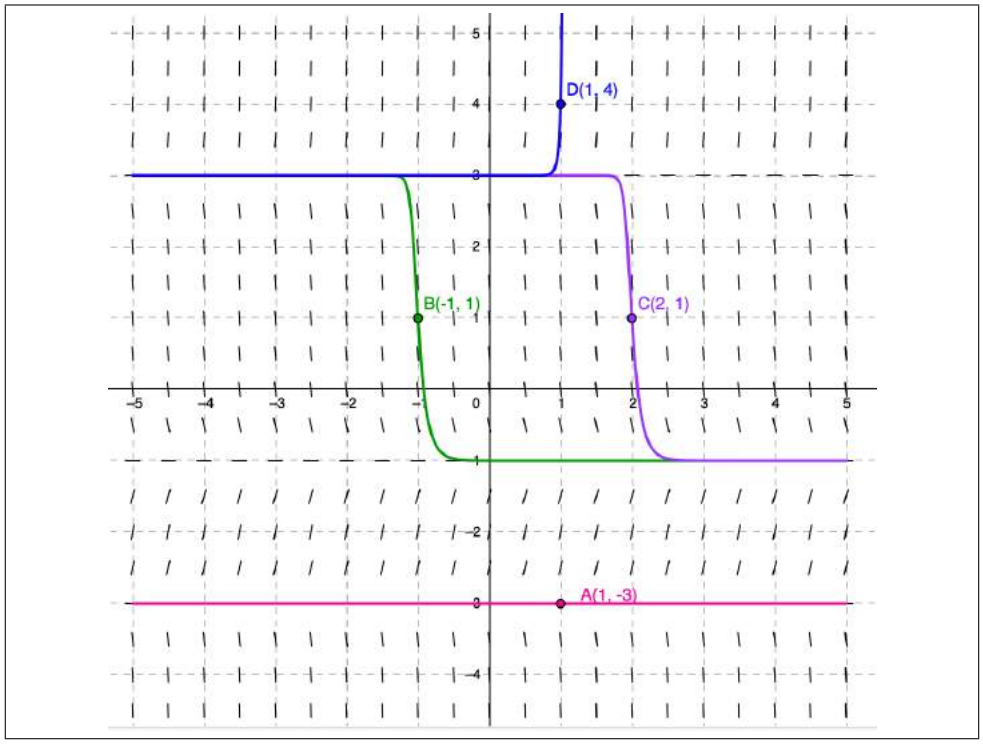


FIGURE 3.4. Direction field for the autonomous equation $y' = (y+1)(y^2-9)$ and several solution curves.

Example 3.7.2 is a rather typical example of an autonomous equation. We make a few remarks about what we see in Example 3.7.2 which hold for all autonomous equations:

Remark 3.7.3. Suppose $y' = f(y)$ is an autonomous equation.

(1) The direction field does not change as you go from left to right, it only changes as you go from bottom to top. This is because the function $f(t, y) = f(y)$ is only a function of y and does not depend on t.
(2) Suppose $y_0(t)$ is a particular solution and $C \in \mathbb{R}$ is a constant. Then $y_0(t + C)$ (a shift of y_0 to the left by C) is also a solution. Indeed:

$$\big(y_0(t+C)\big)' \;=\; y_0'(t+C) \;=\; f\big(y_0(t+C)\big)$$

(3) Suppose $y_0 \in \mathbb{R}$ is such that $f(y_0) = 0$. Then the constant function $y(t) := y_0$ for all t is a solution to $y' = f(y)$. Such a number y_0 is called an **equilibrium point** and the constant function $y(t) := y_0$ is called an **equilibrium solution**.

What about the nonequilibrium solutions? As Example 3.7.2 illustrates, these solutions are strictly increasing/decreasing and will be asymptotic to one of the equilibrium solutions. For this we make the following observations:

(1) Since $y' = f(y)$, if $f(y_0) < 0$, then the solution going through the point (t_0, y_0) will be strictly decreasing.

(2) Likewise, if $f(y_0) > 0$, then the solution going through the point (t_0, y_0) will be strictly increasing.

This qualitative behavior can be succinctly captured by a so-called *phase line*:

Definition 3.7.4. A **phase line** for the equation $y' = f(y)$ is a plot of the y-axis (displayed horizontally) with the following features:

(1) At every equilibrium point y_0 (i.e., where $f(y_0) = 0$), there is a dot.
(2) In a region between two equilibrium points (or between an equilibrium point and $\pm\infty$), if $f(y) < 0$ in that region, then there is an arrow to the left. This tells us that for these y-values, the solution is strictly decreasing.
(3) In a region where $f(y) > 0$, then there is an arrow to the right. This tells us that for these y-values, the solution is strictly increasing.
(4) At each equilibrium point y_0, if the two arrows on either side of y_0 are both pointing towards y_0, then the dot at y_0 is filled in. Otherwise, the dot is not filled in.

Often the phase line is plotted with a vertical $f(y)$-axis as well, superimposed with a graph of the function $f(y)$.

Example 3.7.5. Example of phase line of above example $y' = (y+1)(y^2-9)$. To be included.

There are two types of equilibrium points:

Definition 3.7.6. Consider the autonomous equation $y' = f(y)$. Suppose $y_0 \in \mathbb{R}$ is an equilibrium point (i.e., $f(y_0) = 0$). We say that y_0 is

(1) **asymptotically stable** if a solution which goes through a point $(t_0, y_0 + \epsilon)$, where $|\epsilon| \ll 1$ is very tiny, will asymptotically approach the solution $y(t) = y_0$. These correspond to the filled-in dots on the phase line.
(2) **unstable** if it is not asymptotically stable, i.e., if there is some solution which goes through a point $(t_0, y_0 + \epsilon)$ which "peels off" and is not asymptotic to the solution $y(t) = y_0$. These correspond to the non-filled-in dots on the phase line.

In other words, asymptotically stable equilibrium points act like "sinks", bringing nearby solution curves towards the constant solution at that point. Unstable equilibrium points, at least on one of the two sides, will "repel" nearby solution curves. Since the type of equilibrium point at y_0 is determined by the sign of the function $f(y)$ on both sides of y_0, if we know whether f is strictly increasing/decreasing as it goes through y_0 we can determine its type:

First Derivative Test for Stability 3.7.7. *Suppose y_0 is an equilibrium point for the autonomous equation $y' = f(y)$, and suppose f is differentiable. Then:*

(1) if $f'(y_0) < 0$, then f is strictly decreasing at y_0 and y_0 is asymptotically stable,
(2) if $f'(y_0) > 0$, then f is strictly increasing at y_0 and y_0 is unstable,
(3) if $f'(y_0) = 0$, then no conclusion can drawn and further investigation is needed.

This suggests a general procedure for plotting a direction field with various solution curves:

(1) By studying the function $f(y)$, first construct the phase line, including classifying the equilibrium points as either asymptotically stable or unstable,
(2) In the direction field, plot the equilibrium solutions.
(3) In the other regions, plot solution curves that behave according to the phase line: if the phase line points to the left, the solution should be strictly decreasing and asymptotic to the next lower equilibrium solution (or diverge to $-\infty$). If the phase line points to the right, the solution should be strictly increasing and asymptotic to the next higher equilibrium solution (or diverge to $+\infty$).

CHAPTER 4

Second-order linear differential equations

Recall that an explicit second-order differential equation is an equation of the form

$$y'' \;=\; f(t, y, y')$$

where f is a three-variable function. A solution to this equation is a function $y(t)$ which is at least twice-differentiable such that for every t,

$$y''(t) \;=\; f\big(t, y(t), y'(t)\big)$$

In this chapter, we will study a very special type of second-order differential equation, the so-called *linear* second-order differential equations.

4.1. Overview of second-order linear equations

In general, second-order differential equations (in the fullest generality) are an order of magnitude more complicated than first-order differential equations. For this reason, we will restrict our attention to the simplest type of second-order differential equation, the second-order linear differential equations. As we shall see, there is much we can say about these equations and they have many practical applications.

Definition 4.1.1. A **second-order linear differential equation** is a differential equation which can be put in the form:

$$y''(t) + p(t)y' + q(t)y \;=\; g(t)$$

where the **coefficient functions** p, q, g are functions of the independent variable t only. The function $g(t)$ is referred to as the **forcing term**. If $g(t) = 0$ is the constant zero function, then the differential equation

$$y'' + p(t)y' + q(t)y \;=\; 0$$

is said to be **homogeneous**.

Here is a representative example:

Example 4.1.2 (Simple harmonic motion)**.** Consider the homogeneous second-order linear equation:

$$y'' + \omega^2 y \;=\; 0$$

where $\omega \in \mathbb{R}$ is a constant with $\omega \neq 0$. Consider the functions:

$$y_1(t) \;=\; \cos \omega t \quad \text{and} \quad y_2(t) \;=\; \sin \omega t$$

We claim that these are both solutions (in Section 4.2 we will learn how one finds these solutions). Indeed, note that:

$$\begin{aligned} y_1'(t) &= -\omega \sin \omega t \\ y_2'(t) &= \omega \cos \omega t \\ y_1''(t) &= -\omega^2 \cos \omega t \\ y_2''(t) &= -\omega^2 \sin \omega t \end{aligned}$$

and thus

$$y_1''(t) + \omega^2 y_1(t) = -\omega^2 \cos \omega t + \omega^2 \cos \omega t = 0$$

and

$$y_2''(t) + \omega^2 y_2(t) = -\omega^2 \sin \omega t + \omega^2 \sin \omega t = 0.$$

Are there any other solutions? In this section we will study general properties of the set of solutions to a second-order linear equation. We will *not* learn techniques for actually solving second-order equations in this section, but instead just assume (for the moment) that we have some method of obtaining solutions. Along these lines, the following is relevant:

Existence and Uniqueness Theorem 4.1.3 (Second-Order Linear)**.** *Suppose $p, q, g : I \to \mathbb{R}$ are continuous functions with domain $I \subseteq \mathbb{R}$ an interval. Then given $t_0 \in I$ and any two real numbers $y_0, y_1 \in \mathbb{R}$ there is a unique function $y : I \to \mathbb{R}$ which satisfies the initial value problem:*

(i) $y'' + p(t)y' + q(t) = g(t)$
(ii) $y(t_0) = y_0$ *and* $y'(t_0) = y_1$.

In example 4.1.2 we saw that we had at least two solutions $y_1(t), y_2(t)$ to the equation $y'' + \omega^2 y = 0$. For homogeneous linear equations, given two solutions, we can mass-produce many more solutions.

Definition 4.1.4. Suppose $y_1, y_2 : I \to \mathbb{R}$ are two functions defined on an interval $I \subseteq \mathbb{R}$. A **linear combination** of y_1 and y_2 is any function of the form:

$$C_1 y_1 + C_2 y_2 : I \to \mathbb{R}$$

where $C_1, C_2 \in \mathbb{R}$ are constants.

For example, $3 \cos \omega t - 7 \sin \omega t$ is a linear combination of $\cos \omega t$ and $\sin \omega t$. The following proposition says that the collection of all solutions to a homogeneous second-order linear equation is "closed under linear combinations":

Proposition 4.1.5. *Suppose $y_1(t), y_2(t)$ are solutions to the homogeneous second-order differential equation*

$$y'' + p(t)y' + q(t)y = 0.$$

Then for any $C_1, C_2 \in \mathbb{R}$, the function $C_1 y_1 + C_2 y_2$ is also a solution.

PROOF. Let $C_1, C_2 \in \mathbb{R}$ be arbitrary. Note that

$$\begin{aligned}
& (C_1y_1 + C_2y_2)'' + p(t)(C_1y_1 + C_2y_2)' + q(t)(C_1y_1 + C_2y_2) \\
= \; & (C_1y_1'' + C_2y_2'') + p(t)(C_1y_1' + C_2y_2') + q(t)(C_1y_1 + C_2y_2) \\
& \text{(because the derivative is linear)} \\
= \; & C_1y_1'' + p(t)C_1y_1' + q(t)C_1y_1 + C_2y'' + p(t)C_2y_2' + q(t)C_2y_2 \\
= \; & C_1\big(y_1'' + p(t)y_1' + q(t)y_1\big) + C_2\big(y_2'' + p(t)y_2' + q(t)y_2\big) \\
= \; & C_1 \cdot 0 + C_2 \cdot 0 \\
= \; & 0,
\end{aligned}$$

because y_1 and y_2 both solutions. Thus $C_1y_1 + C_2y_2$ is also a solution. □

When are two solutions "essentially different"? This is captured by the notion of *linear independence*:

Definition 4.1.6. Suppose $y_1, y_2 : I \to \mathbb{R}$ are functions defined on an interval $I \subseteq \mathbb{R}$. We say that y_1 and y_2 are **linearly independent** if: for every $C_1, C_2 \in \mathbb{R}$, if

$$C_1y_1(t) + C_2y_2(t) \;=\; 0 \quad \text{for every } t \in I,$$

then $C_1 = C_2 = 0$. In other words, y_1 and y_2 are linearly independent if the *only* way for a linear combination of y_1 and y_2 to be the constant zero function is with the trivial linear combination $0y_1 + 0y_2$. If y_1 and y_2 are not linearly independent, then we say they are **linearly dependent**.

For two functions y_1 and y_2 to be linearly dependent, this means that either y_1 is a constant multiple of y_2 (i.e., $y_1 = Cy_2$ for some $C \in \mathbb{R}$) or y_2 is a constant multiple of y_1 ($y_2 = Cy_1$ for some $C \in \mathbb{R}$).

Linear independence is ultimately a linear algebra concept and it is one of the most important definitions in undergraduate mathematics.

Example 4.1.7. Here are some examples of pairs of linearly (in)dependent functions:

(1) The functions $y_1 = \cos t$ and $y_2 = \sin t$ are linearly independent.

JUSTIFICATION. Suppose $C_1, C_2 \in \mathbb{R}$ are arbitrary such that

$$C_1 \cos t + C_2 \sin t \;=\; 0 \quad \text{for every } t \in \mathbb{R}. \tag{†}$$

We must show that it *must* be the case that $C_1 = C_2 = 0$. Since (†) holds for *all* $t \in \mathbb{R}$, it holds for $t_0 := 0$. Plugging in this t-value tells us:

$$0 \;=\; C_1 \cos 0 + C_2 \sin 0 \;=\; C_1 \cdot 1 + C_2 \cdot 0 \;=\; C_1$$

and so $C_1 = 0$. Likewise, (†) must also hold for $t_1 := \pi/2$. Plugging in this t-value tells us:

$$0 \;=\; C_1 \cos \pi/2 + C_2 \sin \pi/2 \;=\; C_1 \cdot 0 + C_2 \cdot 1 \;=\; C_2,$$

and so $C_2 = 0$ as well. Since $C_1 = C_2 = 0$, we conclude that $\cos t$ and $\sin t$ are linearly independent. □

(2) The functions e^t and $2e^t$ are *not* linearly independent (i.e., they are linearly *dependent*).

JUSTIFICATION. Note that for $C_1 := 2$ and $C_2 = -1$ we have

$$C_1 e^t + C_2 2e^t \;=\; 2e^t - 2e^t \;=\; 0 \quad \text{for every } t \in \mathbb{R},$$

however C_1 and C_2 are not both zero. □

(3) Suppose $f : \mathbb{R} \to \mathbb{R}$ is any function and $g : \mathbb{R} \to \mathbb{R}$ is the constant zero function ($g(t) := 0$ for all $t \in \mathbb{R}$). Then $f(t)$ and $g(t)$ are linearly dependent.

JUSTIFICATION. Note that for $C_1 := 0$ and $C_2 := 1$ we have

$$C_1 f(t) + C_2 g(t) \;=\; 0 \cdot f(t) + 1 \cdot 0 \;=\; 0 \quad \text{for every } t \in \mathbb{R},$$

although C_1 and C_2 are not both zero (only C_1 is zero, but we need both of them to be zero in order to conclude linear independence). □

As Example 4.1.7 illustrates, it can sometimes be a little tedious to show directly that two functions are linearly independent. Miraculously, for differentiable functions there is a much more systematic way to determine the linear dependence/independence of a pair of functions. This involves computing the so-called *Wronskian*:

Definition 4.1.8. Suppose $u, v : I \to \mathbb{R}$ are two differentiable functions defined on an interval $I \subseteq \mathbb{R}$. Define the **Wronskian** of u and v to be the function $W : I \to \mathbb{R}$ defined by

$$W(t) \;:=\; \det \begin{bmatrix} u(t) & v(t) \\ u'(t) & v'(t) \end{bmatrix} \;:=\; u(t)v'(t) - v(t)u'(t)$$

for all $t \in I$.

You might think that the Wronskian $W(t)$ could in general be any function, but in fact it satisfies the following surprising dichotomy:

Proposition 4.1.9 (Wronskian dichotomy I)**.** *Suppose $p, q, u, v : I \to \mathbb{R}$ are functions defined on an interval $I \subseteq \mathbb{R}$ such that u and v are solutions to*

$$y'' + p(t)y' + q(t)y \;=\; 0$$

Let $W(t)$ be the Wronskian of u and v. Then exactly one of the following two things is true:

(Case 1) $W(t) = 0$ *for* all $t \in I$*, or*
(Case 2) $W(t) \neq 0$ *for* all $t \in I$*.*

PROOF. We are assuming that both u and v satisfy:

$$u'' + pu' + qu \;=\; 0 \quad \text{and} \quad v'' + pv' + qv \;=\; 0.$$

We wish to show that $W = uv' - vu'$ is either everywhere zero, or everywhere nonzero. First, differentiate W:

$$\begin{aligned} W' &= uv'' + u'v' - vu'' - v'u' \\ &= uv'' - vu'' \\ &= u(-pv' - qv) - v(-pu' - qu) \\ &\quad \text{because } u, v \text{ are solutions} \\ &= -puv' - quv + pvu' + quv \\ &= -p(uv' - vu') \\ &= -pW. \end{aligned}$$

Thus, the function $W(t)$ is a solution to the first-order linear homogeneous equation $W' + pW = 0$. Pick t_0 in the domain of W, and suppose $W(t_0) = W_0$. Then by Theorem 3.3.6 we have that

$$W(t) = W_0 \exp\left(-\int_{t_0}^{t} p(s)\, ds\right)$$

Thus, if $W_0 = 0$, we are in Case 1. Otherwise, if $W_0 \neq 0$, we are in Case 2, since the exponential function is never zero. □

Proposition 4.1.9 essentially says that $W(t)$ must be always zero or never zero. It can't be sometimes zero and sometimes not-zero. The dichotomy in Proposition 4.1.9 gives rise to the linear dependence/independence dichotomy:

Proposition 4.1.10 (Wronskian dichotomy II)**.** *Suppose $p, q, u, v : I \to \mathbb{R}$ are functions defined on an interval $I \subseteq \mathbb{R}$ such that u and v are solutions to*

$$y'' + p(t)y' + q(t)y = 0$$

Let $W(t)$ be the Wronskian of u and v. Then:

(Case 1) if there is some $t_0 \in I$ such that $W(t_0) = 0$ (which implies $W(t) = 0$ for all $t \in I$), then u and v are linearly dependent, and

(Case 2) if there is some $t_0 \in I$ such that $W(t_0) \neq 0$ (which implies $W(t) \neq 0$ for all $t \in I$), then u and v are linearly independent.

Proof. **Case 1:** Assume first that we are in Case 1, i.e., there is some $t_0 \in I$ such that $W(t_0) = 0$. Then by Proposition 4.1.9 we know that $W(t) = 0$ for all $t \in I$. We have two subcases.

Case 1(a): Assume that $v(t) = 0$ for every $t \in I$. Then $1 \cdot v(t) + 0 \cdot u(t) = 0$ for every $t \in I$ and so u and v are linearly dependent.

Case 1(b): Assume there is $t_0 \in I$ such that $v(t_0) \neq 0$. By the Bump Lemma 2.2.7, there is $\alpha < t_0 < \beta$ such that $v(t) \neq 0$ for every $t \in (\alpha, \beta) \cap I$. On this interval $(\alpha, \beta) \cap I$, we have

$$\frac{d}{dt}\frac{u}{v} = \frac{u'v - uv'}{v^2} = \frac{-W}{v^2} = 0.$$

Thus by Corollary 2.3.7 there is a constant $C \in \mathbb{R}$ such that $u(t)/v(t) = C$ for every $t \in (\alpha, \beta) \cap I$. I.e., $u(t) = Cv(t)$ for every $t \in (\alpha, \beta) \cap I$. In particular, both $u(t)$ and $Cv(t)$ are solutions to the IVP:

(1) $y'' + py' + qy = 0$
(2) $y(t_0) = u(t_0)$, $y'(t_0) = u'(t_0)$.

By the Existence and Uniqueness Theorem 4.1.3, we conclude that $u(t) = Cv(t)$ for every $t \in I$. Thus $u(t)$ and $v(t)$ are linearly dependent.

Case 2: Suppose there is $t_0 \in I$ such that $W(t_0) \neq 0$. By Proposition 4.1.9 we know that $W(t) \neq 0$ for all $t \in I$. Assume towards a contradiction that $u(t), v(t)$ are linearly dependent. Thus there exists constants $C_1, C_2 \in \mathbb{R}$ such that $(C_1, C_2) \neq (0, 0)$ and that $C_1 u(t) + C_2 v(t) = 0$ for every $t \in I$. This gives us two cases:

Case 2(a): Suppose $C_1 \neq 0$. Then for $C := -C_2/C_1$ we have $u(t) = Cv(t)$ for every $t \in I$. Thus the Wronskian is:

$$W(t) \;=\; uv' - vu' \;=\; Cvv' - v(Cv)' \;=\; 0,$$

a contradiction.

Case 2(b): Suppose $C_2 \neq 0$. This case is similar. □

Example 4.1.11. We return to the first two examples from Example 4.1.7:

(1) Consider the equation:

$$y'' + y \;=\; 0$$

This has solutions $y_1 = \cos t$ and $y_2 = \sin t$. Next we compute the Wronskian:

$$W(t) \;=\; \cos t(\sin t)' - \sin t(\cos t)' \;=\; \cos^2 t + \sin^2 t \;=\; 1.$$

We see that this is everywhere $\neq 0$. Thus by Proposition 4.1.10 we conclude that y_1, y_2 are linearly independent.

(2) Consider the equation:

$$y'' - 2y' + y \;=\; 0$$

We see that $y_1 = e^t$ and $y_2 = 2e^t$ are both solutions. Next we compute the Wronskian:

$$W(t) \;=\; e^t(2e^t)' - 2e^t(e^t)' \;=\; 2e^{2t} - 2e^{2t} \;=\; 0$$

Since $W(t) = 0$ for all t, we conclude by Proposition 4.1.10 that y_1, y_2 are linearly dependent.

We now arrive at the main result of this section:

Theorem 4.1.12. *Suppose y_1, y_2 are linearly independent solutions to the homogeneous second-order linear equation*

$$y'' + p(t)y' + q(t)y \;=\; 0$$

Then the general solution is:

$$y(t; C_1, C_2) \;=\; C_1 y_1(t) + C_2 y_2(t).$$

Proof. Suppose $y : I \to \mathbb{R}$ is an arbitrary solution to $y'' + py' + qy = 0$. We must show there exists constants $C_1, C_2 \in \mathbb{R}$ such that $y = C_1 y_1 + C_2 y_2$. Let $t_0 \in I$. We first must find constants $C_1, C_2 \in \mathbb{R}$ which satisfy:

$$\begin{aligned} C_1 y_1(t_0) + C_2 y_2(t_0) &= y(t_0) \\ C_1 y_1'(t_0) + C_2 y_2'(t_0) &= y'(t_0) \end{aligned}$$

This is possible because y_1, y_2 are assumed to be linearly independent, and thus

$$W(t_0) \;=\; \det \begin{bmatrix} y_1(t_0) & y_2(t_0) \\ y_1'(t_0) & y_2'(t_0) \end{bmatrix} \;\neq\; 0.$$

This implies that the above system has a unique solution. The function $C_1y_1+C_2y_2$ is also a solution to $y''+py'+qy=0$ by Proposition 4.1.5. Furthermore, both $y: I \to \mathbb{R}$ and $C_1y_1+C_2y_2 : I \to \mathbb{R}$ are solutions to the IVP:

(i) $y''+py'+qy=0$
(ii) $y(t_0)=y(t_0)$, $y'(t_0)=y'(t_0)$,

and so by the Existence and Uniqueness Theorem 4.1.3 if follows that $y = C_1y_1 + C_2y_2$ (i.e., these functions $I \to \mathbb{R}$ are equal). This finishes the proof. □

Since a pair of linearly independent solutions to a homogeneous second-order linear equation is capable of producing all other solutions, we call such a pair a *fundamental set of solutions*:

Definition 4.1.13. A **fundamental set of solutions** to the homogeneous second-order equation

$$y''+p(t)y'+q(t)y \;=\; 0$$

is a pair y_1, y_2 of linearly independent solutions. Ultimately, "fundamental set of solutions" refers to the fact that the pair y_1, y_2 satisfies the following two properties:

(1) y_1 and y_2 "generate" all other solutions in the sense that the general solution is $y(t; C_1, C_2) = C_1y_1 + C_2y_2$, and
(2) there is no "redundancy" among y_1 and y_2 (since they are linearly independent), i.e., both solutions are needed to generate all other solutions.

[In linear algebra terms, a "fundamental set of solutions" is a *basis* of the subspace of all solutions.]

Example 4.1.14 (Simple harmonic motion)**.** Find the particular solution to the following initial value problem:

(1) $y''+\omega^2 y=0$ ($\omega \in \mathbb{R}$ a constant, $\omega \neq 0$)
(2) $y(0)=1$, $y'(0)=2$.

We already know that $y_1 = \cos\omega t$ and $y_2 = \sin\omega t$ form a fundamental set of solutions (since $W(t)=\omega \neq 0$). By Theorem 4.1.12, the general solution is:

$$y(t; C_1, C_2) \;=\; C_1\cos\omega t + C_2\sin\omega t$$

for some $C_1, C_2 \in \mathbb{R}$. We will use the initial conditions to solve for C_1, C_2. First note that since $y(0)=1$, we have

$$1 \;=\; y(0) \;=\; C_1\cos 0 + C_2\sin 0 \;=\; C_1.$$

Second, taking a derivative of the general solution yields:

$$y'(t) \;=\; -C_1\omega\sin\omega t + C_2\omega\cos\omega t$$

and the condition $y'(0)=2$ gives

$$2 \;=\; y'(0) \;=\; -C_1\omega\sin 0 + C_2\omega\cos 0 \;=\; C_2\omega$$

and so $C_2 \;=\; 2/\omega$. Thus the particular solution is

$$y(t) \;=\; \cos\omega t + \frac{2}{\omega}\sin\omega t.$$

Note: really we obtained a system of equations:

$$\begin{aligned} 1\cdot C_1 + 0\cdot C_2 &= 1 \\ 0\cdot C_1 + \omega\cdot C_2 &= 2 \end{aligned}$$

This particular system is immediate to solve, but in general it might be more complicated and require Gaussian Elimination (or whatever your favorite method of solving a 2×2 system is).

We end by answering a question which is implicit in the above discussion:

Question 4.1.15. *Given $p, q : I \to \mathbb{R}$ defined on an interval $I \subseteq \mathbb{R}$, does there always exist two linearly independent solutions $y_1, y_2 : I \to \mathbb{R}$ of the homogeneous second-order linear differential equation:*

$$y'' + p(t)y' + q(t)y = 0.$$

ANSWER. Yes! By the Existence Uniqueness Theorem 4.1.3, we can arbitrarily choose $t_0 \in I$ and then obtain two solutions $y_1, y_2 : I \to \mathbb{R}$ which satisfy the initial conditions:

(1) $y_1(t_0) = 1$, $y_1'(t_0) = 0$
(2) $y_2(t_0) = 0$, $y_2'(t_0) = 1$.

We claim that y_1, y_2 are linearly independent. Indeed, note that:

$$W(t_0) = \det \begin{bmatrix} y_1(t_0) & y_2(t_0) \\ y_1'(t_0) & y_2'(t_0) \end{bmatrix} = \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1.$$

Thus the Wronskian of y_1, y_2 is nonzero at least at the value t_0. By Proposition 4.1.10 it follows that $W(t) \neq 0$ for every $t \in I$ and also that y_1, y_2 are linearly independent. □

4.2. Homogeneous second-order linear equations with constant coefficients

In this section we study a very special case of homogeneous second-order linear equations, those with *constant coefficients*:

$$y'' + py' + qy = 0$$

where $p, q \in \mathbb{R}$ are constant functions. The *simple harmonic motion* equation (Example 4.1.2) is already an example of such an equation. To study these equations, we need to introduce an auxiliary device, the so-called *characteristic polynomial*:

Definition 4.2.1. The **characteristic polynomial** associated to the homogeneous second-order linear equation

$$y'' + py' + qy = 0$$

(where $p, q \in \mathbb{R}$ are constant functions) is the quadratic polynomial

$$f(\lambda) = \lambda^2 + p\lambda + q$$

in the variable λ. A root of the characteristic polynomial is called a **characteristic root**.

Recall that the *quadratic formula* gives us the roots of a quadratic equation:

$$\lambda_1, \lambda_2 = \frac{-p \pm \sqrt{p^2 - 4q}}{2}$$

Furthermore, the nature of the two roots λ_1, λ_2 fall into three cases, depending on the value of the *discriminant* $p^2 - 4q$:

(1) If $p^2 - 4q > 0$, then $\lambda_1 \neq \lambda_2$ are distinct and both real numbers.
(2) If $p^2 - 4q = 0$, then $\lambda_1 = \lambda_2$ are the same real number.

(3) If $p^2 - 4q < 0$, then $\lambda_1 \neq \lambda_2$ are distinct but they are not real numbers (they are complex numbers).

We shall study these three cases separately.

Distinct real roots. In this subsection, we fix a homogeneous second-order linear differential equation with constant coefficients:

$$y'' + py' + qy = 0$$

and we let

$$f(\lambda) = \lambda^2 + p\lambda + q$$

be its characteristic polynomial. Furthermore, we assume that f has two distinct real roots λ_1 and λ_2.

Theorem 4.2.2 (Distinct real roots)**.** *The general solution to*

$$y'' + py' + qy = 0$$

when $\lambda_1 \neq \lambda_2$ are distinct and real is:

$$y(t; C_1, C_2) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}.$$

PROOF. We first claim that $e^{\lambda_i t}$ is a solution, for $i = 1, 2$. Note that

$$\begin{aligned}
(e^{\lambda_i t})'' + p(e^{\lambda_i t}) + qe^{\lambda_i t} &= \lambda_i^2 e^{\lambda_i t} + p\lambda_i e^{\lambda_i t} + qe^{\lambda_i t} \\
&= (\lambda_i^2 + p\lambda_i + q)e^{\lambda_i t} \\
&= f(\lambda_i)e^{\lambda_i t} \\
&= 0e^{\lambda_i t} \quad \text{because } \lambda_i \text{ is a root of } f(\lambda) \\
&= 0.
\end{aligned}$$

Thus both $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$ are solutions. Next, we claim they are linearly independent. Indeed, note that:

$$\begin{aligned}
W(t) &= \det \begin{bmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ (e^{\lambda_1 t})' & (e^{\lambda_2 t})' \end{bmatrix} \\
&= \det \begin{bmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ \lambda_1 e^{\lambda_1 t} & \lambda_2 e^{\lambda_2 t} \end{bmatrix} \\
&= \lambda_2 e^{\lambda_1 t} e^{\lambda_2 t} - \lambda_2 e^{\lambda_1 t} e^{\lambda_2 t} \\
&= (\lambda_2 - \lambda_1) e^{(\lambda_1 + \lambda_2)t} \\
&\neq 0
\end{aligned}$$

since $\lambda_2 - \lambda_1 \neq 0$ and $e^{(\lambda_1+\lambda_2)t}$ is never zero. Thus by Theorem 4.1.12 we conclude that $e^{\lambda_1 t}, e^{\lambda_2 t}$ is a fundamental set of solutions and that

$$y(t; C_1, C_2) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}.$$

is the general solution. □

Example 4.2.3. We will solve the IVP:

(i) $y'' - 3y' + 2y = 0$
(ii) $y(0) = 2, y'(0) = 1$.

First we compute the zeros of the characteristic polynomial:

$$f(\lambda) \;=\; \lambda^2 - 3\lambda + 2$$

By the quadratic formula, the two zeros are

$$\lambda_1, \lambda_2 \;=\; \frac{3 \pm \sqrt{9-8}}{2} \;=\; 2, 1$$

Thus by Theorem 4.2.2 the general solution is

$$y(t; C_1, C_2) \;=\; C_1 e^{2t} + C_2 e^t.$$

Now we need to use the initial condition to find the values of C_1, C_2. First note that

$$2 \;=\; y(0) \;=\; C_1 e^{2\cdot 0} + C_2 e^0 \;=\; C_1 + C_2$$

and since $y'(t) = 2C_1 e^{4t} + C_2 e^t$, we also have

$$1 \;=\; y'(0) \;=\; 2C_1 e^{2\cdot 0} + C_2 e^0 \;=\; 2C_1 + 1C_2$$

Thus we have a system of equations:

$$\begin{aligned} C_1 + C_2 &= 2 \\ 2C_1 + C_2 &= 1 \end{aligned}$$

There are many ways to solve this, one way is Gaussian Elimination:

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 1 & 1 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 3 \end{array}\right]$$

Thus $C_1 = -1$ and $C_2 = 3$. We conclude that the particular solution to the IVP is:

$$y(t) \;=\; -e^{2t} + 3e^t.$$

Repeated real roots. In this subsection, we study the situation where

$$y'' + py' + qy \;=\; 0$$

has repeated characteristic roots $\lambda_1 = \lambda_2$. Note that the proof of Theorem 4.2.2 above already shows that $e^{\lambda_1 t}$ is a solution. This begs the following question:

Question 4.2.4. *How do we find a second linearly independent solution to $y'' + py' + qy = 0$?*

Answer. Let $y_1(t) = e^{\lambda_1 t}$ be the first known solution. Since λ_1 is a double root of $f(\lambda) = \lambda^2 + p\lambda + q$, it follows that $f(\lambda) = (\lambda - \lambda_1)^2 = \lambda^2 - 2\lambda_1\lambda + \lambda_1^2$. Thus $p = -2\lambda_1$, i.e., $\lambda_1 = -p/2$, and $q = \lambda_1^2 = p^2/4$.

We shall guess that a second solution is of the form $y_2(t) = v(t)e^{\lambda_1 t}$, where $v(t)$ is an unknown function. We shall determine $v(t)$. First we compute:

$$\begin{aligned} y_2' &= e^{\lambda_1 t}(v' + \lambda_1 v) \\ y_2'' &= e^{\lambda_1 t}(v'' + 2\lambda_1 v' + \lambda_1^2 v) \end{aligned}$$

Thus, in order for y_2 to be a solution, we need the following to be equal to zero:

$$\begin{aligned} y_2'' + py_2' + qy_2 &= e^{\lambda_1 t}(v'' + 2\lambda_1 v' + \lambda_1^2 v) + pe^{\lambda_1 t}(v' + \lambda_1 v) + qve^{\lambda_1 t} \\ &= e^{\lambda_1 t}(v'' + 2\lambda_1 v' + \lambda_1^2 v + p(v' + \lambda_1 v) + q) \\ &= e^{\lambda_1 t}\big(v'' - pv' + p^2 v/4 + p(v' - pv/2) + p^2 v/4\big) \\ &= e^{\lambda_1 t} v''. \end{aligned}$$

Since $e^{\lambda_1 t} \neq 0$ for all t, we require that $v'' = 0$. In this case, we get $v = At + B$ is linear, and so we can take a second solution of the form $y_2(t) = (At + B)e^{\lambda_1 t}$. Since we require that $y_2(t)$ is linearly independent with $y_1(t)$, it suffices to take $y_2(t) = te^{\lambda_1 t}$. □

We summarize the above as follows:

Theorem 4.2.5 (Repeated real roots)**.** *The general solution to*

$$y'' + py' + qy \;=\; 0$$

when $\lambda_1 = \lambda_2$ are not distinct (and real) is:

$$y(t; C_1, C_2) \;=\; C_1 e^{\lambda_1 t} + C_2 t e^{\lambda_1 t}.$$

Proof. This is a worksheet exercise. □

Example 4.2.6. We will solve the following IVP:

(1) $y'' - 2y' + y = 0$
(2) $y(0) = 2, y'(0) = -1$

First we consider the characteristic polynomial:

$$f(\lambda) \;=\; \lambda^2 - 2\lambda + \lambda \;=\; (\lambda - 1)^2.$$

We see that $\lambda_1 = \lambda_2 = 1$ is a repeated root. Thus by Theorem 4.2.5 the general solution is

$$y(t; C_1, C_2) \;=\; C_1 e^t + C_2 t e^t$$

Now we need to use our initial condition to solve for C_1, C_2. First note that

$$2 \;=\; y(0) \;=\; C_1 e^0 + C_2 \cdot 0 \cdot e^0 \;=\; C_1.$$

Next we differentiate our general solution:

$$y'(t) \;=\; (C_1 + C_2)e^t + C_2 t e^t$$

to get

$$-1 \;=\; y'(0) \;=\; C_1 + C_2.$$

This yields the system

$$\begin{aligned} C_1 &= 2 \\ C_1 + C_2 &= -1 \end{aligned}$$

We see that $C_1 = 2$, $C_2 = -3$. Thus our particular solution is

$$y(t) \;=\; 2e^t - 3te^t.$$

Complex (non-real) roots. We finally consider the case where $\lambda_1 \neq \lambda_2 \in \mathbb{C}$, i.e., the case when the discriminant $p^2 - 4q < 0$ of the characteristic polynomial is negative which yields two distinct complex (non-real) roots. First we briefly recall some fact about the *complex numbers*:

(1) A **complex number** is a number of the form $z = a + bi$, where $a, b \in \mathbb{R}$ and $i^2 = -1$ is the **imaginary unit**. We denote the set of all complex numbers by $\mathbb{C}$.
(2) Given a complex number $z = a+bi$, we define its **real part** to be $\mathrm{Re}(z) := a$ and its **imaginary part** to be $\mathrm{Im}(z) := b$.
(3) Given a complex number $z = a + bi$, we define its **complex conjugate** to be $\overline{z} := a - bi$.

(4) Here are some facts about the complex conjugate of a complex number $z = a + bi$:
   (a) $\overline{\overline{z}} = z$
   (b) $\mathrm{Re}(z) = (z + \overline{z})/2$
   (c) $\mathrm{Im}(z) = (z - \overline{z})/2i$
   (d) $z = \overline{z}$ iff $z \in \mathbb{R}$ iff $b = 0$.
   (e) for $w \in \mathbb{C}$ we have $\overline{z + w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z} \cdot \overline{w}$

(5) The complex exponential function behaves according to **Euler's formula**:

$$e^{a+bi} \;=\; e^a(\cos b + i \sin b)$$

(6) Suppose $f(\lambda) = \lambda^2 + p\lambda + q$ is a polynomial with real coefficients $p, q \in \mathbb{R}$ and a complex (non-real) root $\lambda_1 = a + bi$. Then $\lambda_2 := \overline{\lambda_1} = a - bi$ is also a complex root, i.e., the complex roots of a real polynomial occur in *complex conjugate pairs*.

(7) Suppose $z(t)$ is a complex-valued function such that $z(t) = x(t) + y(t)i$, where $x(t), y(t)$ are real valued functions. Then

$$\frac{d}{dt}z(t) \;=\; \frac{d}{dt}x(t) + i\frac{d}{dt}y(t)$$

i.e., complex-valued functions can be differentiated by separately differentiating the real and imaginary parts in the usual way.

If we allow ourselves to consider complex-valued functions as a solution to our differential equation, then we obtain the following analogue of the distinct real-roots case (Theorem 4.2.2):

Theorem 4.2.7 (Distinct complex roots, complex version)**.** *The general solution to*

$$y'' + py' + qy \;=\; 0$$

when $\lambda_1 = a + bi, \lambda_2 = a - bi$ are distinct and complex is:

$$y(t; C_1, C_2) \;=\; C_1 e^{(a+bi)t} + C_2 e^{(a-bi)t} \;=\; C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$$

Proof. The proof is the same as the proof of Theorem 4.2.2. □

Of course, ultimately we are interested in real-valued functions as solutions. For this, the following is useful:

Observation 4.2.8. *Suppose $z(t)$ is a complex-valued function which is a solution to*

$$y'' + py' + qy \;=\; 0,$$

where $p, q \in \mathbb{R}$. Then:

(1) the function $\overline{z(t)}$ is also a solution.

Furthermore, since the set of all solutions is closed under linear combinations, it follows that:

(2) $\mathrm{Re}\big(z(t)\big)$ is a real-valued solution, and
(3) $\mathrm{Im}\big(z(t)\big)$ is a real-valued solution.

This observation and Euler's formula yield the following:

Theorem 4.2.9 (Distinct complex roots, real version)**.** *The general solution to*

$$y'' + py' + qy \;=\; 0$$

when $\lambda_1 = a + bi, \lambda_2 = a - bi$ *are distinct and complex is:*

$$y(t; C_1, C_2) \;=\; C_1 e^{at}\cos bt + C_2 e^{at}\sin bt.$$

PROOF. This is a worksheet exercise. □

Example 4.2.10. We will solve the following IVP:

(1) $y'' + 2y' + 2y = 0$
(2) $y(0) = 2, y'(0) = 3.$

First we consider the characteristic polynomial:

$$f(\lambda) \;=\; \lambda^2 + 2\lambda + 2$$

By the quadratic formula, we see that the characteristic roots are:

$$\lambda_1, \lambda_2 \;=\; \frac{-2 \pm \sqrt{4-8}}{2} \;=\; -1 \pm i$$

Thus $\lambda_1 = a + bi = -1 + i$, where $a = -1$ and $b = 1$. By Theorem 4.2.9 the general solution is

$$y(t; C_1, C_2) \;=\; C_1 e^{-t}\cos t + C_2 e^{-t}\sin t$$

Next we use our initial condition to solve for C_1, C_2. Note that

$$2 \;=\; y(0) \;=\; C_1$$

Then we differentiate the general solution:

$$y'(t) \;=\; -e^{-t}(C_1\cos t + C_2\sin t) + e^{-t}(-C_1\sin t + C_2\cos t)$$

to get

$$3 \;=\; y'(0) \;=\; -C_1 + C_2.$$

This yields the system

$$\begin{aligned} C_1 &= 2 \\ -C_1 + C_2 &= 3 \end{aligned}$$

and so $C_1 = 2$, $C_2 = 5$. We conclude that our particular solution is

$$y(t) \;=\; 2e^{-t}\cos t + 5e^{-t}\sin t.$$

4.3. The method of undetermined coefficients

In this section we discuss a method for solving *inhomogeneous* second-order linear equations. The method is called *the method of undetermined coefficients* which also is sometimes called *the method of (judicious) guessing.* This method does not always work, but it works for a large enough class of differential equations that it is worth discussing. The first order of business is to discuss *inhomogeneous* equations in general.

Inhomogeneous equations. Recall that we are ultimately interested in second-order linear differential equations of the form

$$y'' + p(t)y' + q(t)y \;=\; g(t)$$

When the *forcing term* $g(t) = 0$ for all t, then the differential equation is *homogeneous*; otherwise, it is *inhomogeneous*. We have already studied the structure of the general solution to a homogeneous equation in Section 4.1, and we have seen how to solve homogeneous equations with constant coefficients in Section 4.2. The following theorem tells us how to form the general solution of an inhomogeneous solution *provided* that we know the general solution of the corresponding homogeneous equation *and* we are somehow able to obtain at least one particular solution to the inhomogeneous equation:

Theorem 4.3.1 (General solution to inhomogeneous equation)**.** *Suppose $y_p(t)$ is a particular solution to the inhomogeneous equation*

$$y'' + p(t)y' + q(t)y \;=\; g(t) \tag{A}$$

and that $y_1(t), y_2(t)$ form a fundamental set of solutions to the corresponding homogeneous equation

$$y'' + p(t)y' + q(t)y \;=\; 0. \tag{B}$$

Then the general solution to the inhomogeneous equation (A) is

$$y(t) \;=\; y(t; C_1, C_2) \;=\; C_1y_1(t) + C_2y_2(t) + y_p(t).$$

Proof. We need to show two things. First we will show that for any choice of $C_1, C_2 \in \mathbb{R}$, $y(t)$ is indeed a solution to (A). Note that:

$$\begin{aligned}
&\big(y(t)\big)'' + p(t)\big(y(t)\big)' + q(t)y(t) \\
=\;& \big(C_1y_1(t) + C_2y_2(t) + y_p(t)\big)'' \\
&+ p(t)\big(C_1y_1(t) + C_2y_2(t) + y_p(t)\big)' \\
&+ q(t)\big(C_1y_1(t) + C_2y_2(t) + y_p(t)\big) \\
=\;& \big(C_1y_1 + C_2y_2\big)'' + p(t)\big(C_1y_1 + C_2y_2\big)' + q(t)\big(C_1y_1 + C_2y_2\big) \\
&+ y_p'' + p(t)y_p' + q(t)y_p \\
&\quad \text{because the derivative is linear} \\
=\;& 0 + y_p'' + p(t)y_p' + q(t)y_p \\
&\quad \text{because } C_1y_1 + C_2y_2 \text{ is a solution to (B)} \\
=\;& g(t) \quad \text{because } y_p \text{ is a solution to (A).}
\end{aligned}$$

Thus $y(t) = C_1y_1 + C_2y_2 + y_p$ is a solution to (A).

Next we will show that an arbitrary solution $y(t)$ of (A) *must* be of the form $y(t) = C_1y_1 + C_2y_2 + y_p$ for some choice of $C_1, C_2 \in \mathbb{R}$. Consider the function

$\tilde{y}(t) := y(t) - y_p(t)$. Note that

$$\begin{aligned}
&\tilde{y}'' + p(t)\tilde{y}' + q(t)\tilde{y} \\
&= (y - y_p)'' + p(t)(y - y_p)' + q(t)(y - y_p) \\
&= y'' + p(t)y' + q(t)y - y_p'' - p(t)y_p' - q(t)y_p \\
&\qquad \text{because the derivative is linear} \\
&= g(t) - g(t) \\
&\qquad \text{because both } y \text{ and } y_p \text{ are solution to (A)} \\
&= 0.
\end{aligned}$$

Thus $\tilde{y}(t)$ is a solution to (B). Since $y_1(t), y_2(t)$ form a fundamental set of solutions to (B), there are constants $C_1, C_2 \in \mathbb{R}$ such that $\tilde{y} = C_1 y_1 + C_2 y_2$. Thus $y(t) - y_p(t) = C_1 y_1 + C_2 y_2$ and thus $y(t) = C_1 y_1 + C_2 y_2 + y_p$. □

In other words, to find the general solution to an inhomogeneous solution

$$y'' + p(t)y' + q(t)y \;=\; g(t),$$

you need to do the following:

(1) First, find a fundamental set of solutions y_1, y_2 to the homogeneous equation $y'' + p(t)y' + q(t)y = 0$ (possibly using techniques from Section 4.2 if p and q are constants).
(2) Second, find *one* particular solution y_p to the inhomogeneous equation $y'' + p(t)y' + q(t)y = g(t)$ (possibly using the method of undetermined coefficients below if p and q are constants, or the method of variation of parameters from Section 4.4).
(3) Third, write down the general solution:
$$y(t) \;=\; C_1 y_1(t) + C_2 y_2(t) + y_p(t)$$
(4) (If necessary) Fourth, if you are solving an IVP, then use the initial conditions to solve for the precise values of C_1, C_2 from the general solution in the same way you would solve an IVP for a homogeneous equation.

Finally, we remark that Theorem 4.3.1 (and its proof) ultimately belongs to the subject of linear algebra (when viewed appropriately). Note that the only relevant feature from differential equations that got used in the proof was that the LHS is linear as a result of the derivative being linear (i.e., $(f + g)' = f' + g'$). We will revisit this theme of "general solution to inhomogeneous is general solution of homogeneous plus particular solution of inhomogeneous" in the next chapter.

Method of undetermined coefficients. We now introduce the method of undetermined coefficients. This method allows us to find particular solutions of an inhomogeneous second-order linear differential equation

$$y'' + p(t)y' + q(t)y \;=\; g(t)$$

provided:

(1) p and q are constant functions, and
(2) $g(t)$ is a "nice enough" function.

Ultimately, the *method of undetermined coefficients* involves *guessing* a so-called **trial solution**, and then plugging in that trial solution to determine a specific particular solution. We illustrate this first with an example:

Example 4.3.2. Find a particular solution to:

$$y'' + 3y' + 2y \;=\; 4e^{-3t}.$$

SOLUTION. Here the forcing term is $g(t) = 4e^{-3t}$. We will *guess* that there is a particular solution of the form $y_p(t) = ae^{-3t}$, where $a \in \mathbb{R}$ is an *undetermined coefficient* (i.e., an unknown coefficient we need to somehow determine). Thus in this case our "trial solution" is a function $y_p(t) = ae^{-et}$ To find a, we plug the trial solution $y_p(t)$ into the equation:

$$y_p'' + 3y_p' + 2y_p \;=\; 9ae^{-3t} - 9ae^{-3t} + 2ae^{-3t} \;=\; 4e^{-3t}.$$

This simplifies to

$$(9a - 9a + 2a)e^{-3t} \;=\; 2ae^{-3t} \;=\; 4e^{-3t}$$

and so $2a = 4$, i.e., $a = 2$. Thus the function $y_p(t) = 2e^{-3t}$ is a particular solution to $y'' + 3y' + 2y = 4e^{-3t}$. □

How did we know to guess the trial solution ae^{-3t} in the above example? For many cases, the trial solution can be correctly guessed by using the following heuristics:

(1) The trial solution should include the function $g(t)$ as a special case. In the above example, $g(t) = 4e^{-3t}$ is also of the form ae^{-3t}.
(2) The trial solution should be a family of functions "closed under the derivative." In the above example, the derivative of a function of the form "ae^{-3t}" is $-3ae^{-et}$ which is also of the form "ae^{-3t}" (where "$-3a$" plays the role of "a").

In practice you can just look up the trial solution you are supposed to guess according to:

Method of Undetermined Coefficients 4.3.3. *Suppose $y'' + py' + qy = g(t)$ is an inhomogeneous differential equation such that:*

(a) $p, q \in \mathbb{R}$ are constants, and
(b) $g(t)$ is not *a solution to the homogeneous solution $y'' + py' + qy = 0$.*

Then the following gives the trial solution you should guess depending on the form of the forcing function $g(t)$ (where $A, B, a, b, r, \omega \in \mathbb{R}$, $P(t)$ is a polynomial and $p_0(t), p_1(t)$ are polynomials of the same degree as P). If the forcing function $g(t)$ is of the form...

(1) e^{rt}, then the trial solution is $y_p(t) = ae^{rt}$.
(2) $A\cos\omega t + B\sin\omega t$, then the trial solution is $y_p(t) = a\cos\omega t + b\sin\omega t$.
(3) $P(t)$, then the trial solution is $y_p(t) = p_0(t)$.
(4) $P(t)\cos\omega t$ or $P(t)\sin\omega t$, then the trial solution is

$$y_p(t) \;=\; p_0(t)\cos\omega t + p_1(t)\sin\omega t.$$

(5) $e^{rt}\cos\omega t$ or $e^{rt}\sin\omega t$, then the trial solution is

$$y_p(t) \;=\; e^{rt}(a\cos\omega t + b\sin\omega t).$$

(6) $e^{rt}P(t)\cos\omega t$ or $e^{rt}P(t)\sin\omega t$, then the trial solution is

$$y_p(t) \;=\; e^{rt}\big(p_0(t)\cos\omega t + p_1(t)\sin\omega t\big).$$

If $g(t)$ is a solution to $y'' + py' + qy$, then use the trial solution $ty_p(t)$, and if that does not work, then use the trial solution $t^2y_p(t)$.

Here is an example which falls in case (2) in 4.3.3

Example 4.3.4. Find a particular solution to

$$y'' + 4y = \cos 3t.$$

SOLUTION. Since $g(t) = \cos 3t$, the Method of Undetermined Coefficients 4.3.3 tells us our trial solution should be $y_p(t) = a\cos 3t + b\sin 3t$, where $a, b \in \mathbb{R}$ are undetermined coefficients we need to determine. First, we need to compute y_p' and y_p'':

$$\begin{aligned} y_p'(t) &= -3a\sin 3t + 3b\cos 3t \\ y_p''(t) &= -9a\cos 3t - 9b\sin 3t \end{aligned}$$

Plugging this into the LHS of the differential equation yields:

$$y_p'' + 4y_p = -9a\cos 3t - 9b\sin 3t + 4(a\cos 3t + b\sin 3t) = -5a\cos 3t - 5b\sin 3t.$$

This needs to equal $\cos 3t$, so we get:

$$-5a\cos 3t - 5b\sin 3t = \cos 3t = 1\cos 3t + 0\sin 3t.$$

This yields the system:

$$\begin{aligned} -5a &= 1 \\ -5b &= 0 \end{aligned}$$

so we find that $a = -1/5$ and $b = 0$. Thus we find that a particular solution is:

$$y_p(t) = -\frac{1}{5}\cos 3t. \qquad \square$$

Here is an example which falls into case (3) of 4.3.3:

Example 4.3.5. Find a particular solution to

$$y'' + 6y' + 8y = 2t - 3.$$

SOLUTION. Since $g(t) = 2t - 3$ is a polynomial of degree 2, the Method of Undetermined Coefficients 4.3.3 tells us our trial solution should be $y_p(t) = a_1 t + a_0$ (a polynomial of the same degree as $g(t)$), where $a_1, a_0 \in \mathbb{R}$ are undetermined coefficients which we need to determine. First we compute y_p' and y_p'':

$$\begin{aligned} y_p'(t) &= a_1 \\ y_p''(t) &= 0 \end{aligned}$$

Next we plug $y_p(t)$ into the LHS of the differential equation to get:

$$y_p'' + 5y_p' + 8y_p = 0 + 5a_1 + 8(a_1 t + a_0) = 8a_1 t + (5a_1 + 8a_0)$$

which needs to equal the RHS $2t - 3$, which gives:

$$8a_1 t + (5a_1 + 8a_0) = 2t - 3.$$

This yields the system:

$$\begin{aligned} 8a_1 &= 2 \\ 5a_1 + 8a_0 &= -3 \end{aligned}$$

We can solve this using Gaussian Elimination:

$$\left[\begin{array}{cc|c} 8 & 0 & 2 \\ 5 & 8 & -3 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c} 1 & 0 & 1/4 \\ 0 & 1 & -17/32 \end{array}\right]$$

Thus a particular solution is

$$y_p(t) \;=\; \frac{8t-17}{32}. \qquad \square$$

The following *superposition principle* shows how to handle forcing terms which are a linear combination of forcing terms covered in 4.3.3:

Superposition Principle 4.3.6. *Suppose $y_f(t)$ is a particular solution to*

$$y'' + p(t)y' + q(t)y \;=\; f(t)$$

and $y_g(t)$ is a particular solution to

$$y'' + p(t)y' + q(t)y \;=\; g(t).$$

Then for $\alpha, \beta \in \mathbb{R}$, the function $y(t) := \alpha y_f(t) + \beta y_g(t)$ is a solution to

$$y'' + p(t)y' + q(t)y \;=\; \alpha f(t) + \beta g(t).$$

Here is an example of the Superposition Principle in use:

Example 4.3.7. Find a particular solution to

$$y'' + 2y' + 2y \;=\; 2 + \cos 2t.$$

SOLUTION. We have two forcing terms here, $f(t) := 2$ and $g(t) := \cos 2t$. We need to handle each one separately.

First we will find a particular solution $y_f(t)$ to

$$y'' + 2y' + 2y \;=\; 2.$$

Since $f(t) = 2$ is a degree 0 polynomial, the trial solution is $y_f(t) = a_0$, also a degree 0 polynomial. Plugging this in to the LHS and equating this to the RHS yields:

$$y_f'' + 2y_f' + 2y_f \;=\; 2a_0 \;=\; 2.$$

Thus we find that $a_0 = 1$ and thus $y_f(t) = 1$ is a particular solution.

Next we will find a particular solution $y_g(t)$ to

$$y'' + 2y' + 2y \;=\; \cos 2t.$$

Since $g(t) = \cos 2t$, the trial solution is $y_g(t) = a\cos 2t + b\sin 2t$. First note that

$$\begin{aligned} y_g'(t) &= -2a\sin 2t + 2b\cos 2t \\ y_g''(t) &= -4a\cos 2t - 4b\sin 2t \end{aligned}$$

Plugging this into the LHS and equating it to the RHS yields:

$$\begin{aligned} y_g''+2y_g'+2y_g &= (-4a\cos 2t - 4b\sin 2t) + 2(-2a\sin 2t + 2b\cos 2t) + 2(a\cos 2t + b\sin 2t) \\ &= (-2a+4b)\cos 2t + (-2b-4a)\sin 2t \;=\; \cos 2t \end{aligned}$$

This yields the system:

$$\begin{aligned} -2a + 4b &= 1 \\ -4a - 2b &= 0 \end{aligned}$$

We can find a, b by Gaussian Elimination:

$$\left[\begin{array}{cc|c} -2 & -4 & 1 \\ -4 & -2 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c} 1 & 0 & 1/6 \\ 0 & 1 & -1/3 \end{array}\right]$$

Thus we find that $a = 1/6, b = -1/3$, and so

$$y_g(t) \;=\; \frac{\cos 2t - 2\sin 2t}{6}.$$

We conclude that a particular solution to the original differential equation is:

$$y_p(t) \;=\; y_f(t) + y_g(t) \;= 1 + \frac{\cos 2t - 2\sin 2t}{6}. \qquad \square$$

4.4. Variation of parameters

In this section we introduce a method of finding a particular solution to an inhomogeneous equation

$$y'' + p(t)y' + q(t)y \;=\; g(t)$$

provided we already know a fundamental set of solutions $y_1(t), y_2(t)$ to the associated homogeneous equation:

$$y'' + p(t)y' + q(t)y \;=\; 0.$$

The method essentially will rely on the following fact about 2×2 systems of equations (which will be justified in the next chapter):

Fact 4.4.1. Suppose $a, b, c, d, e, f \in \mathbb{R}$ are numbers such that

$$W \;:=\; \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;=\; ad - bc \;\neq\; 0.$$

Then the system

$$\begin{aligned} ax + by \;&=\; e \\ cx + dy \;&=\; f \end{aligned}$$

has the unique solution:

$$x \;=\; \frac{de - bf}{W}, \quad y \;=\; \frac{-ce + af}{W}.$$

Variation of Parameters 4.4.2. *Suppose $y_1(t), y_2(t)$ is a fundamental set of solutions to:*

$$y'' + p(t)y' + q(t)y \;=\; 0,$$

(in particular, $W(t) := y_1y_2' - y_2y_1' \neq 0$ for all t). Then the inhomogeneous equation:

$$y'' + p(t)y' + q(t)y \;=\; g(t)$$

has the following as a particular solution:

$$y_p(t) \;=\; y_1 \int \frac{-y_2(t)g(t)\,dt}{W(t)} + y_2 \int \frac{y_1(t)g(t)\,dt}{W(t)}.$$

Proof. We know that

$$y(t) \;=\; C_1y_1(t) + C_2y_2(t)$$

is the general solution to the homogeneous equation. The idea is to replace the constants C_1, C_2 with unknown functions v_1v_2 and look for a particular solution to the inhomogeneous equation of the form

$$y_p \;=\; v_1y_1 + v_2y_2.$$

First we compute the first derivative of y_p:

$$y_p' \;=\; v_1y_1' + v_1'y_1 + v_2y_2' + v_2'y_2 \;=\; (v_1y_1' + v_2y_2') + (v_1'y_1 + v_2'y_2)$$

Ideally, we do not want to deal with any second-order derivatives of v_1 and v_2, since otherwise we would not be making our lives any easier. Furthermore, in some sense requiring y_p to be a solution to the inhomogeneous equation places only one condition on the two unknown functions v_1, v_2, thus we have some "freedom" to impose a second condition in case it helps. Thus, we will additionally assume:

$$v_1'y_1 + v_2'y_2 \;=\; 0. \tag{A}$$

Now we compute the second derivative of y_p:

$$y_p'' \;=\; \big[(v_1y_1' + v_2y_2') + \underbrace{(v_1'y_1 + v_2'y_2)}_{=0}\big]' \;=\; v_1y_1'' + v_1'y_1' + v_2y_2'' + v_2'y_2'$$

Next we plug y_p into the LHS of the differential equation and simplify. Note that:

$$\begin{aligned} y_p'' + py_p' + qy_p \;&=\; v_1y_1'' + v_1'y_1' + v_2y_2'' + v_2'y_2' + p(v_1y_1' + v_2y_2') + q(v_1y_1 + v_2y_2) \\ &=\; v_1(y_1'' + py_1' + qy_1) + v_2(y_2'' + py_2' + qy_2) \\ &\quad\; + v_1'y_1' + v_2'y_2' \\ &=\; v_1'y_1' + v_2'y_2', \end{aligned}$$

because y_1, y_2 are solutions to the homogeneous equation. Setting LHS equal to RHS yields:

$$v_1'y_1' + v_2'y_2' \;=\; g(t). \tag{B}$$

Now, we combine (A) and (B) into a single system in the unknown "variables" v_1', v_2':

$$\begin{aligned} y_1v_1' + y_2v_2' \;&=\; 0 \\ y_1'v_1' + y_2'v_2' \;&=\; g(t) \end{aligned}$$

Since y_1, y_2 is a fundamental set of solutions, we see that

$$W(t) \;=\; \det\begin{bmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{bmatrix} \;\neq\; 0$$

for every t. Thus by Fact 4.4.1, we get

$$v_1' \;=\; \frac{-y_2(t)g(t)}{W(t)} \quad\text{and}\quad v_2' \;=\; \frac{y_1(t)g(t)}{W(t)}.$$

Finally, v_1, v_2 are obtained by integrating v_1', v_2' with respect to t. □

Example 4.4.3. Find a particular solution to the inhomogeneous equation

$$y'' + y \;=\; \tan t$$

on the interval $(-\pi/2, \pi/2)$.

SOLUTION. First we find a fundamental set of solutions to $y''+y=0$. Note that the characteristic polynomial is $f(\lambda) = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$. Thus $\lambda_1, \lambda_2 = \pm i$, and so a fundamental set of solutions is $y_1(t) = \cos t$, $y_2(t) = \sin t$. Next we compute the Wronskian:

$$W(t) \;=\; \det\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} \;=\; \cos t(\cos t) - \sin t(-\sin t) \;=\; \cos^2 t + \sin^2 t \;=\; 1.$$

Next, we get v_1:

$$\begin{aligned} v_1(t) &= \int \frac{-y_2(t)g(t)\,dt}{W(t)} \\ &= \int -\sin t \tan t\,dt \\ &= -\int \frac{\sin^2 t}{\cos t} dt \\ &= -\int \frac{\cos^2 t - 1}{\cos t} dt \\ &= \sin t - \ln|\sec t + \tan t| \\ &= \sin t - \ln(\sec t + \tan t) \end{aligned}$$

since $\sec t + \tan t \geq 0$ on the interval $(-\pi/2, \pi/2)$. Next we get v_2:

$$\begin{aligned} v_2(t) &= \int \frac{y_1(t)g(t)\,dt}{W(t)} \\ &= \int \cos t \tan t\,dt \\ &= \int \sin t \\ &= -\cos t. \end{aligned}$$

We conclude that a particular solution is:

$$y_p(t) = y_1v_1 + y_2v_2 = \cos t\big(\sin t - \ln(\sec t + \tan t)\big) + \sin t(-\cos t)$$

$$= -\cos t \ln(\sec t + \tan t). \qquad \square$$

Example 4.4.4. Find a particular solution to:

$$t^2y'' + ty' - y = t\ln t,$$

where $y_1(t) = t$ and $y_2(t) = 1/t$ is a fundamental set of solutions to

$$t^2y'' + ty' - y = 0.$$

PROOF. In order to use Variation of Parameters, the coefficient of y'' in the inhomogeneous equation needs to equal 1. Thus we will divide the equation through by t^2 to obtain:

$$y'' + \frac{y'}{t} - \frac{y}{t^2} = \frac{\ln t}{t}.$$

Thus $g(t) = (\ln t)/t$. Furthermore, since the differential equation only makes sense on the interval $(0, +\infty)$, this is where we will work. Next, we need to compute the Wronskian of the two fundamental solutions:

$$W(t) = \det\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} = t(-1/t^2) - (1/t)\cdot 1 = -\frac{1}{t} - \frac{1}{t} = -\frac{2}{t}.$$

Next we compute $v_1(t)$:

$$\begin{aligned} v_1(t) &= \int \frac{-y_2(t)g(t)\,dt}{W(t)} \\ &= \int \frac{-(1/t)(\ln t)/t}{-2/t}dt \\ &= \frac{1}{2}\int \frac{\ln t}{t}\,dt \\ &= \frac{(\ln t)^2}{4}. \end{aligned}$$

We also compute $v_2(t)$:

$$\begin{aligned} v_2(t) &= \int \frac{y_1(t)g(t)\,dt}{W(t)} \\ &= \int \frac{t(\ln t)/t\,dt}{-2/t} \\ &= -\frac{1}{2}\int t\ln t\,dt \\ &= -\frac{t^2(2\ln t-1)}{8}. \end{aligned}$$

We conclude that a particular solution is:

$$y_p(t) = y_1v_1 + y_2v_2 = \frac{t(\ln t)^2}{4} - \frac{t(2\ln t-1)}{8} \qquad \square$$

CHAPTER 5

Linear algebra II

We have already seen in Chapter 1 that the device of *augmented matrix* is very useful for systematically solving systems of equations. For the next step in our linear algebra journey, we will treat matrices as a fundamental object of interest in their own right and work with them almost exclusively.

5.1. Matrices and vectors

Definition 5.1.1. Suppose $m, n \geq 1$. A **matrix (of size $m \times n$)** is a rectangular array of real numbers with m rows and n columns:

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

Sometimes we abbreviate a matrix by writing:

$$A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \quad \text{or just} \quad A = (a_{ij})$$

if the size of the matrix A is clear from context. Given $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, the number a_{ij} is called the (i, j)**-entry (or component) of** A. We denote the set of all $m \times n$ matrices (with real numbers as entries) by $\mathrm{Mat}_{m \times n}(\mathbb{R})$.

A matrix in $\mathrm{Mat}_{m \times 1}(\mathbb{R})$ with only one column:

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

is often called a **column vector**. We will often denote $\mathrm{Mat}_{m \times 1}$ by $\mathbb{R}^m$, and write column vectors with bold letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{z}$, etc.

For each $m, n \geq 1$, we define the **zero matrix** in $\mathrm{Mat}_{m \times n}(\mathbb{R})$ to be the $m \times n$ matrix where every entry is $= 0$:

$$0_{m \times n} := \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Sometimes we will denote $0_{m \times n}$ as just 0 when it is clear from context that we are talking about the $m \times n$ zero matrix (and not, for instance, the number $0 \in \mathbb{R}$).

A *matrix* is, in a certain sense, a vast generalization of a *number*. Just as we can add, subtract, multiply, and divide numbers, we can *sometimes* do versions of these things with matrices. Here are the most fundamental operations defined for matrices:

Definition 5.1.2. Fix $m, n \geq 1$. Given two matrices $A, B \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, we define their **matrix sum** $A + B \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ to be the $m \times n$ matrix whose (i, j)-entry is $a_{ij} + b_{ij}$, i.e.,

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

Furthermore, given $\alpha \in \mathbb{R}$, we define the **scalar multiple** of A by α to be the matrix $\alpha A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ whose (i, j)-entry is $\alpha a_{i,j}$, i.e.,

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

Example 5.1.3.

(1) Here is an example of how matrix addition works (for matrices in $\mathrm{Mat}_{3\times 2}(\mathbb{R})$):

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \\ 10 & 14 \end{bmatrix}$$

(2) Here is an example of how scalar multiplication works (for column vectors in $\mathbb{R}^4$, which is the same thing as matrices in $\mathrm{Mat}_{4\times 1}(\mathbb{R})$):

$$3 \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \\ -3 \end{bmatrix}$$

Fact 5.1.4. Suppose $m, n \geq 1$, $A, B, C \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, and $\alpha, \beta \in \mathbb{R}$. Then the following facts[1] about matrix addition and scalar multiplication hold:

(1) $(A + B) + C = A + (B + C)$ (associativity of addition)
(2) $0_{m\times n} + A = A + 0_{m\times n} = A$ (additive identity)
(3) $A + (-1)A = 0_{m\times n}$ (additive inverse)
(4) $A + B = B + A$ (commutativity of addition)
(5) $\alpha(A + B) = \alpha A + \alpha B$ (right distributivity)
(6) $(\alpha + \beta)A = \alpha A + \beta A$ (left distributivity)

[1]These facts say that the set $\mathrm{Mat}_{m\times n}(\mathbb{R})$ equipped with matrix addition and scalar multiplication is a **vector space** over the real numbers $\mathbb{R}$.

(7) $(\alpha\beta)A = \alpha(\beta A)$ (associativity of scalar multiplication)
(8) $1 \cdot A = A$ (here $1 \in \mathbb{R}$ is a scalar)

Definition 5.1.5. Suppose $n \geq 1$. A **linear combination** of column vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ is an expression of the form:

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_m\mathbf{v}_m,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$ are scalars.

5.2. Matrix equations

Matrices and vectors give us a superior way of writing and talking about system of equations:

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

In order to make sense of this, the first step is to define the product of a matrix with a column vector:

Definition 5.2.1. Suppose $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^n$. We define the **product** to be the column vector $A\mathbf{x} \in \mathbb{R}^m$ whose $(i,1)$-entry is

$$(A\mathbf{x})_{i,1} = \sum_{k=1}^{n} A_{i,k}x_k$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Another way to say this: we can write the matrix A as a collection of n column vectors in $\mathbb{R}^m$:

$$A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{bmatrix}$$

Then the product $A\mathbf{x}$ is defined to be the linear combination:

$$A\mathbf{x} := x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Written yet another way, this is:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

Here is an example of a product of a matrix with a column vector:

Example 5.2.2. Consider the matrix and column vector:

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 7 & -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$$

Then the product $A\mathbf{x}$ is:

$$A\mathbf{x} = \begin{bmatrix} 1 & 3 & 5 \\ 7 & -2 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1(-1) + 3(2) + 3(5) \\ 7(-1) + 2(-2) + 4(3) \end{bmatrix} = \begin{bmatrix} 20 \\ 1 \end{bmatrix}$$

Warning 5.2.3. In order for the product of a matrix A and a column vector $\mathbf{x}$ to be defined and make sense, the number of columns of A needs to equal the number of rows of $\mathbf{x}$. Otherwise, the product $A\mathbf{x}$ is not defined and thus does not make sense. For example, you can multiply a 2×2 matrix with a 2×1 column vector, but you cannot multiply a 2×3 matrix with a 2×1 column vector.

Here are some basic facts about matrix multiplication which we will use:

Fact 5.2.4. Suppose $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$. Then:

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

Systems of equations. Now, we can interpret a system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{5.1}$$

as a matrix equation:

(1) First, we define the **coefficient matrix** to be the $m \times n$ matrix $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ defined by:

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

(2) Second, we combine our unknown variables $x_1, \ldots, x_n$ into a single **vector of unknowns $\mathbf{x}$** (of size $n \times 1$):

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

(3) Third, we combine our right-hand side parameters $b_1, \ldots, b_m$ into a single column vector $\mathbf{b} \in \mathbb{R}^m$:

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

(4) Finally, we can translate the system (5.1) into the **matrix equation**:

$$A\mathbf{x} = \mathbf{b} \tag{5.2}$$

If we write $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}$ in terms of its column vectors, then we can also express the equation (5.2) as:

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \;=\; \mathbf{b}.$$

We could also express (5.2) by writing everything out fully:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

There are advantages and disadvantages to each choice of notations, although ultimately these are just equivalent ways of rewriting (5.1) in terms of matrices and vectors.

Remark 5.2.5. When it comes to solving matrix equations $A\mathbf{x} = \mathbf{b}$, everything from Chapter 1 applies. For instance, suppose we wish to solve the matrix equation:

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

To solve this, we set up the corresponding augmented matrix and take it to RREF:

$$\begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 1 \end{bmatrix} \xrightarrow{R_2 - 3R_1 \to R_2} \begin{bmatrix} 1 & 2 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$
$$\xrightarrow{-\frac{1}{2}R_2 \to R_2} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1/2 \end{bmatrix}$$
$$\xrightarrow{R_1 - 2R_2 \to R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1/2 \end{bmatrix}$$

and we see that $(x_1, x_2) = (1, -1/2)$ is the unique solution to the system of equation. In other words, the column vector

$$\mathbf{x} \;=\; \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$$

is the unique solution to the matrix equation

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

Indeed, note that the product

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix} \;=\; \begin{bmatrix} 1(1) + 2(-1/2) \\ 3(1) + 4(-1/2) \end{bmatrix} \;=\; \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

gives the correct right-hand side of the equation.

5.3. Nullspace, linear independence, and dimension

In this section we will dive deeper into important features of a matrix equation:

$$A\mathbf{x} \;=\; \mathbf{b}$$

Ultimately, we will be establishing important definitions and basic properties involving these definitions, in order to better understand how the solutions to a matrix equation look and behave. Since we already learned how to completely solve matrix equations (disguised as systems of equations) in Chapter 1, there will not be any

new computational methods in this section, however, we will repurpose the method of Gaussian Elimination to answer many more types of questions related to matrix equations and their solutions.

There will be a strong analogy between the nature of solutions to a matrix equation and the nature of solutions to a linear differential equation (since both are secretly applications of abstract linear algebra). The first similarity already shows up in the following definition:

Definition 5.3.1. Suppose $A \in \text{Mat}_{m\times n}(\mathbb{R})$ and consider the matrix equation

$$A\mathbf{x} \;=\; \mathbf{b} \tag{5.3}$$

where $\mathbf{b} \in \mathbb{R}^m$. We say that the equation (5.3) is **homogeneous** if $\mathbf{b} = 0_{m\times 1}$ is the zero vector in $\mathbb{R}^m$. Otherwise, if $\mathbf{b} \neq 0$, then we say that the equation (5.3) is **inhomogeneous**.

We will first be interested in studying homogeneous matrix equations. In this context, the following is a very important definition:

Definition 5.3.2. Suppose $A \in \text{Mat}_{m\times n}(\mathbb{R})$. We define the **nullspace** of A to be the following subset of $\mathbb{R}^n$:

$$\text{null}(A) \;:=\; \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\} \;\subseteq\; \mathbb{R}^n$$

In other words, the nullspace $\text{null}(A)$ of the matrix A is the set of all solutions to the homogeneous equation $A\mathbf{x} = 0$.

From Chapter 1 we already know how to compute the nullspace of a matrix:

Example 5.3.3. Find the nullspace of the following matrix:

$$A \;=\; \begin{bmatrix} 1 & 1 & 0 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Proof. We need to find the set of all vectors $\mathbf{x} \in \mathbb{R}^4$ such that $A\mathbf{x} = 0$. This means the same thing as finding all solutions to the system of equations:

$$\begin{aligned} x_1 + x_2 + 4x_4 \;&=\; 0 \\ x_3 + 2x_4 \;&=\; 0. \end{aligned}$$

To do this, we set up the system as an augmented matrix and take it to RREF:

$$\left[\begin{array}{cccc|c} 1 & 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \end{array}\right]$$

Here we see that the augmented matrix is already in RREF, so we can read off the solutions. We see that x_2, x_4 are free variables, so the general solution is:

$$\begin{aligned} x_1 \;&=\; -x_2 - 4x_4 \\ x_2 \;&=\; x_2 \\ x_3 \;&=\; -2x_4 \\ x_4 \;&=\; x_4 \end{aligned}$$

Which we can write in parametric form as a set of linear combination of $\mathbb{R}^4$-vectors:

$$\text{null}(A) \;=\; \left\{ x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \end{bmatrix} : x_2, x_4 \in \mathbb{R} \right\} \qquad \square$$

The nullspace of a matrix is also closed under linear combinations (analogous to Proposition 4.1.5):

Proposition 5.3.4. *Suppose $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$. Let $\mathbf{x}_0, \mathbf{x}_1 \in \mathrm{null}(A)$, $\alpha \in \mathbb{R}$ be arbitrary. Then:*

(1) $0 \in \mathrm{null}(A)$, where $0 = 0_{n\times 1}$ is the zero vector in $\mathbb{R}^n$,
(2) $\mathbf{x} + \mathbf{y} \in \mathrm{null}(A)$, and
(3) $\alpha\mathbf{x} \in \mathrm{null}(A)$.

[In linear algebra terms, this says that $\mathrm{null}(A)$ is a subspace of $\mathbb{R}^n$.]

Proof. (1) Let $0 = 0_{n\times 1}$ be the zero vector in $\mathbb{R}^n$. Then by Definition 5.2.1 it follows that $A0_{n\times 1} = 0_{m\times 1}$. Thus $0_{n\times 1} \in \mathrm{null}(A)$.

(2) Note that

$$\begin{aligned} A(\mathbf{x}+\mathbf{y}) &= A\mathbf{x} + A\mathbf{y} \\ &= 0_{m\times 1} + 0_{m\times 1} \quad \text{since } \mathbf{x}, \mathbf{y} \in \mathrm{null}(A) \\ &= 0_{m\times 1} \end{aligned}$$

and thus $\mathbf{x} + \mathbf{y} \in \mathrm{null}(A)$.

(3) Note that

$$\begin{aligned} A(\alpha\mathbf{x}) &= \alpha A\mathbf{x} \\ &= \alpha 0_{m\times 1} \quad \text{since } \mathbf{x} \in \mathrm{null}(A) \\ &= 0_{m\times 1}. \end{aligned}$$

Thus $\alpha\mathbf{x} \in \mathrm{null}(A)$. □

Next, we want to say a few words about how to efficiently describe a nullspace. We now define a notation which allows us to describe a large set of vectors in $\mathbb{R}^n$:

Definition 5.3.5. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$. The **span** of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is the set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_k$:

$$\mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k) := \{\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k : \alpha_1, \ldots, \alpha_k \in \mathbb{R}\}$$

In other words, the span of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is the set of all vectors which can be "created" from $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

Example 5.3.6. Here are some common usages of *span*:

(1) We can describe $\mathbb{R}^2$ as a span, in multiple different ways:

$$\mathbb{R}^2 = \mathrm{span}\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right) = \mathrm{span}\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right) = \mathrm{span}\left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right)$$

(2) We can describe $\mathbb{R}^3$ as a span:

$$\mathbb{R}^3 = \mathrm{span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right)$$

(There are infinitely many other ways to describe $\mathbb{R}^3$ as a span).

(3) Returning to Example 5.3.3 above, we found that

$$\mathrm{null}(A) = \left\{ x_2 \begin{bmatrix}-1\\1\\0\\0\end{bmatrix} + x_4 \begin{bmatrix}-4\\0\\-2\\1\end{bmatrix} : x_2, x_4 \in \mathbb{R} \right\}$$

Another way of writing this in terms of span:

$$\text{null}(A) \;=\; \text{span}\left(\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-4\\0\\-2\\1\end{bmatrix}\right)$$

In some sense, it is better to express:

$$\mathbb{R}^2 \;=\; \text{span}\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right) \quad \text{or} \quad \mathbb{R}^2 \;=\; \text{span}\left(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right)$$

instead of

$$\mathbb{R}^2 \;=\; \text{span}\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right)$$

The reason is because in this last description, two of the four vectors are redundant. For instance, the third and fourth can already be written as linear combinations of the first and second vectors, and vice-versa. The next concept we will introduce is "non-redundancy", better known as *linear independence* (compare to Definition 4.1.6, also recall our earlier statement: the notion of *linear independence* is one of the most important definitions in undergraduate mathematics):

Definition 5.3.7. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$. We say that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are **linearly independent** if for every $c_1, \ldots, c_k \in \mathbb{R}$, if $c_1\mathbf{x}_1 + \cdots c_k\mathbf{x}_k = 0$, then $c_1 = c_2 = \cdots = c_k = 0$. In other words, $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent iff the homogeneous matrix equation

$$A\mathbf{c} \;=\; 0 \quad \text{where} \quad A = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x_2} & \cdots & \mathbf{x_k}\end{bmatrix}$$

has exactly one solution, $\mathbf{c} = 0_{n\times 1}$.

Otherwise, we say that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are **linearly dependent**. In other words, $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent if there exists $c_1, \ldots, x_k \in \mathbb{R}^n$ such that $c_i \neq 0$ for at least one $i \in \{1, \ldots, k\}$. In this case, the linear combination $c_1\mathbf{x}_1 + \cdots c_k\mathbf{x}_k = 0$ is called a **nontrivial dependence relation**.

We can use Gaussian Elimination to check if a collection of vectors is linearly (in)dependent:

Example 5.3.8. Here is an example of linear independence and linear dependence.

(1) The vectors

$$\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\4\\8\end{bmatrix}, \begin{bmatrix}3\\9\\27\end{bmatrix}$$

are linearly independent. Why? We need to show that the the equation $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0_{3\times 1}$ has only one solution $(c_1, c_2, c_3) = (0, 0, 0)$. This is equivalent to showing that the system of equations:

$$\begin{aligned} c_1 + 2c_3 + 3c_3 &= 0 \\ 2c_1 + 4c_2 + 9c_3 &= 0 \\ 3c_1 + 8c_2 + 27c_3 &= 0 \end{aligned}$$

has a unique solution. To see this, we set up an augmented matrix and take it to RREF:

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 2 & 4 & 9 & 0 \\ 3 & 8 & 27 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Since every variable is a pivot variable and there are no free variables, we see that there is a unique solution, which must be $(c_1, c_2, c_3) = (0, 0, 0)$.

(2) The vectors

$$\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix}$$

are linearly dependent. Why? We need to find a nontrivial dependence relation between these three vectors which is equivalent to finding a nontrivial solution to the following system of equations:

$$\begin{aligned} 3c_2 - 3c_3 &= 0 \\ c_1 - c_2 + 3c_3 &= 0 \\ 2c_1 + 4c_2 &= 0 \end{aligned}$$

To do this, we set up an augmented matrix and take it to RREF:

$$\left[\begin{array}{ccc|c} 0 & 3 & -3 & 0 \\ 1 & -1 & 3 & 0 \\ 2 & 4 & 0 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We see that c_3 is a free variable and so the general solution is:

$$\begin{aligned} c_1 &= -2c_3 \\ c_2 &= c_3 \\ c_3 &= c_3 \end{aligned}$$

which we can write in parametric form:

$$\left\{ c_3 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} : c_3 \in \mathbb{R} \right\}$$

To get a nontrivial solution, we can choose, for instance, $c_3 := 1$ to get the solution $(c_1, c_2, c_3) = (-2, 1, 1)$. This gives us a nontrivial dependence relation:

$$(-2) \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + 1 \cdot \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} + 1 \cdot \begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We conclude these three vectors are linearly dependent.

Remark 5.3.9. (1) The empty set $\emptyset$ of vectors in, say, $\mathbb{R}^n$ is considered to be linearly independent. This corresponds to $k = 0$ in Definition 5.3.7.

(2) Suppose $k = 1$ in Definition 5.3.7. This means we have a set of one vector $\mathbf{x}_1 \in \mathbb{R}^n$. Then $\mathbf{x}_1$ is linearly independent iff $\mathbf{x}_1 \neq 0_{n\times 1}$. This is because the linear combination $c_1\mathbf{x}_1 = 0_{m\times 1}$ requires either $c_1 = 0$ or $\mathbf{x}_1 = 0_{m\times 1}$ in order to be true. If $\mathbf{x}_1 \neq 0_{m\times 1}$, then necessarily $c_1 = 0$.

(3) For two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, $\mathbf{x}_1, \mathbf{x}_2$ are linearly dependent iff there exists $\alpha \in \mathbb{R}$ such that either $\mathbf{x}_1 = \alpha\mathbf{x}_2$ or $\mathbf{x}_2 = \alpha\mathbf{x}_1$.

(4) For three or more vectors, linear dependence does *not* mean "one vector is a constant multiple of one of the others". For instance, the follow three vectors in $\mathbb{R}^2$:
$$\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}$$
are linearly dependent because we have a nontrivial dependence relation:
$$1 \cdot \begin{bmatrix}1\\0\end{bmatrix} + 1 \cdot \begin{bmatrix}0\\1\end{bmatrix} + (-1) \cdot \begin{bmatrix}1\\1\end{bmatrix} \;=\; \begin{bmatrix}0\\0\end{bmatrix},$$
even though none of the three vectors is exactly a multiple of the other two.

(5) If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$ are linearly *independent*, then for $\ell \leq k$, the smaller collection $\mathbf{x}_1, \ldots, \mathbf{x}_\ell$ is automatically linearly independent.

(6) If $\mathbf{x}_1 \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$ are linearly *dependent*, then any larger collection
$$\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_{k+\ell}$$
is automatically linearly dependent.

(7) If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are k distinct vectors in $\mathbb{R}^n$, and $n < k$, then necessarily $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent. This is because the corresponding homogeneous matrix equation
$$A\mathbf{c} \;=\; 0_{n\times 1} \quad \text{where} \quad A \;=\; \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k\end{bmatrix}$$
corresponds to an $n \times k$ matrix A with more columns than rows, so there is guaranteed to be at least one free variable (hence infinitely many solutions).

Combining the notions of *span* and *linear independence*, we arrive at the notion of *basis* and *dimension* (of a nullspace):

Definition 5.3.10. Suppose $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$. A **basis** of $\mathrm{null}(A)$ is a collection of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ such that:

(1) $\mathrm{null}(A) = \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ (so $\mathbf{x}_1, \ldots, \mathbf{x}_k$ can make all of $\mathrm{null}(A)$ by linear combinations), and

(2) $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent (so none of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are unnecessary or redundant).

We define the **dimension** of $\mathrm{null}(A)$ to be the number of vectors in a basis of $\mathrm{null}(A)$. Thus
$$\dim \mathrm{null}(A) \;:=\; k \quad \Longleftrightarrow \quad \text{there is a basis } \mathbf{x}_1, \ldots, \mathbf{x}_k \text{ of } \mathrm{null}(A) \text{ with } k \text{ vectors}$$

Example 5.3.11. Returning to Example 5.3.3, we already saw that:
$$\mathrm{null}(A) \;=\; \mathrm{span}\left(\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-4\\0\\-2\\1\end{bmatrix}\right)$$
Since the vectors
$$\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-4\\0\\-2\\1\end{bmatrix}$$

are linearly independent, we conclude that

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

is a basis of $\text{null}(A)$ and thus $\dim \text{null}(A) = 2$.

Here are some general facts to know, which we state without proof:

Fact 5.3.12. Suppose $A \in \text{Mat}_{m\times n}(\mathbb{R})$

(1) In general, $\text{null}(A)$ will have infinitely many possible bases, but all of these bases have the same size. Thus the definition of $\dim \text{null}(A)$ does not depend on a particular choice of basis.
(2) Recall that the **rank** of A (denoted $\text{rank}(A)$) is the number of pivots in the RREF of A. In general, $\dim \text{null}(A)$ is equal to the number of free variables in the RREF of A. Since the number of pivot variables plus the number of free variables, this yields the important **rank-nullity formula**:
$$\text{rank}(A) + \dim \text{null}(A) \ = \ n \ = \ \#\text{ of columns of } A.$$
(3) A basis for $\text{null}(A)$ can be obtained by solving the homogeneous equation $A\mathbf{c} = 0_{m\times 1}$ in the usual way with Gaussian Elimination, writing the solutions in parametric form with the free variables as parameters, then collecting each vector which gets multiplied by a free variable. This (finite) collection of vectors will be a basis for $\text{null}(A)$.

Finally, we have the following fact for inhomogeneous equations (analogous to Theorem 4.3.1):

Proposition 5.3.13. *Suppose $A \in \text{Mat}_{m\times n}(\mathbb{R})$ and $\mathbf{b} \in \mathbb{R}^m$, and assume $\mathbf{b} \neq 0_{m\times 1}$. Consider the inhomogeneous equation:*

$$(\dagger) \qquad A\mathbf{x} \ = \ \mathbf{b}$$

Suppose we have one particular solution $\mathbf{x}_p \in \mathbb{R}^n$ to (†). Then the set of all solutions to (†) is:

$$\{\mathbf{x}_p + \mathbf{x}_h : \mathbf{x_h} \in \text{null}(A)\}$$

In other words, every solution to (†) is equal to our particular solution $\mathbf{x}_p$ plus a solution $\mathbf{x}_h$ to the homogeneous solution $A\mathbf{x} = 0_{m\times 1}$.

Proof. Let $\mathbf{x}_h \in \text{null}(A)$ and $\mathbf{x}_p$ be our particular solution to (†). Note that:

$$\begin{aligned} A(\mathbf{x}_p + \mathbf{x}_h) &= A\mathbf{x}_p + A\mathbf{x}_h \\ &= \mathbf{b} + 0_{m\times 1} \\ &= \mathbf{b}. \end{aligned}$$

Thus $\mathbf{x}_p + \mathbf{x}_h$ is also a solution to (†). Conversely, suppose $\mathbf{x}_i$ is an arbitrary solution to (†). Note that

$$\begin{aligned} A(\mathbf{x}_i - \mathbf{x}_p) &= A\mathbf{x}_i - A\mathbf{x}_p \\ &= \mathbf{b} - \mathbf{b} \\ &= 0_{m\times 1}. \end{aligned}$$

Thus $\mathbf{x}_i - \mathbf{x}_p \in \text{null}(A)$, so there is $\mathbf{x}_h \in \text{null}(A)$ such that $\mathbf{x}_i - \mathbf{x}_p = \mathbf{x}_h$. Thus $\mathbf{x}_i = \mathbf{x}_p + \mathbf{x}_h$. □

Here are some more facts about the number of solutions of a matrix equation in terms of the terminology from this section:

Fact 5.3.14. Suppose $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and $\mathbf{b} \in \mathbb{R}^m$.

(1) The following are equivalent:
   (a) there does not exist any solutions to $A\mathbf{x} = \mathbf{b}$,
   (b) the system corresponding to $A\mathbf{x} = \mathbf{b}$ is inconsistent,
   (c) there does not exist a particular solution $\mathbf{x}_p$ to $A\mathbf{x} = \mathbf{b}$.

We define the matrix equation $A\mathbf{x} = \mathbf{b}$ to be **inconsistent** if any of the equivalent conditions of (1) above. We say that $A\mathbf{x} = \mathbf{b}$ is **consistent** otherwise.

(2) Suppose $A\mathbf{x} = \mathbf{b}$ is consistent. The following are equivalent:
   (a) there is a unique solution to $A\mathbf{x} = \mathbf{b}$,
   (b) there is a unique solution to the system corresponding to $A\mathbf{x} = \mathbf{b}$,
   (c) $\mathrm{null}(A) = \{0_{n\times 1}\}$,
   (d) $\dim \mathrm{null}(A) = 0$,
   (e) there are no free variables,
   (f) every variable is a pivot variable,
   (g) $\mathrm{rank}(A) = n$.

(3) Suppose $A\mathbf{x} = \mathbf{b}$ is consistent. The following are equivalent:
   (a) there are infinitely many solutions to $A\mathbf{x} = \mathbf{b}$,
   (b) there are infinitely many solutions to the system corresponding to $A\mathbf{x} = \mathbf{b}$,
   (c) $\mathrm{null}(A) \neq \{0_{n\times 1}\}$,
   (d) $\dim \mathrm{null}(A) \geq 1$,
   (e) there is at least one free variable,
   (f) $\mathrm{rank}(A) < n$.

Thus, the distinction between 1 solution versus infinitely many solutions to $A\mathbf{x} = \mathbf{b}$ is entirely determined by $\mathrm{null}(A)$.

5.4. Square matrices and determinants

In anticipation of Chapter 6, in this section we take a closer look at *square matrices*.

Definition 5.4.1. We call a matrix A a **square matrix** if $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ for some $n \geq 1$.

Example 5.4.2. (1) Here are some square matrices of various sizes:

$$\begin{bmatrix}1\end{bmatrix}, \quad \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}, \quad \begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{bmatrix}, \quad \begin{bmatrix}1 & 2 & 3 & 4\\ 5 & 6 & 7 & 8\\ 9 & 10 & 11 & 12\\ 13 & 14 & 15 & 16\end{bmatrix}$$

(2) Suppose $n \geq 1$. We define the **identity matrix** to be the square matrix $I = I_{n\times n} \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ which has 1's on the main diagonal and 0's in all other entries, i.e.,

$$(I)_{i,j} = \begin{cases}1 & \text{if } i = j\\ 0 & \text{if } i \neq j\end{cases}$$

Written out, the identity matrix looks like:

$$I_{n\times n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

The special property of the identity matrix is that for every $\mathbf{v} \in \mathbb{R}^n$, $I_{n\times n}\mathbf{v} = \mathbf{v}$, i.e., multiplying a column vector by an appropriately-sized identity matrix always returns the original vector.

We are interested in answering the following question about square matrices:

Question 5.4.3. *Given a square matrix $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, how can we tell if $\mathrm{null}(A) \neq \{0\}$? Put another way, how can we tell if there exists $\mathbf{v} \in \mathbb{R}^n$ such that $A\mathbf{v} = 0$ but $\mathbf{v} \neq 0$?*

Of course, one way to answer Question 5.4.3 is to just compute a basis for $\mathrm{null}(A)$, and then check whether the basis is empty or nonempty. However for our purposes this method is way too cumbersome (in the next section, we will ask this question for an infinite family of matrices simultaneously). Fortunately, there is a much easier way to answer this question: with *determinants*.

Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ is a square matrix. Then associated to A is a number $\det(A) \in \mathbb{R}$ called the **determinant of** A. In other words, there is a function:

$$\det : \mathrm{Mat}_{n\times n}(\mathbb{R}) \to \mathbb{R}$$

We will not carefully define this function, but we will give the formula for how to compute it. For this class, ultimately we will treat the determinant as a black-box and take on faith all of its relevant properties.

Computing the determinant. For $n = 1$, computing the determinant is easy:

$$\text{Given } A = \begin{bmatrix} a_{11} \end{bmatrix} \in \mathrm{Mat}_{1\times 1}(\mathbb{R}), \text{ we have } \det A = a_{11}.$$

For $n = 2$, there is also a fairly simple formula for computing the determinant:

$$\text{Given } A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathrm{Mat}_{2\times 2}(\mathbb{R}), \text{ we have } \det A = a_{11}a_{22} - a_{21}a_{12}.$$

Now suppose $n \geq 2$, and let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. Then for any $i, j \in \{1, \ldots, n\}$ we define the **ij-cofactor matrix of** A to be the matrix $\widetilde{A}_{ij} \in \mathrm{Mat}_{(n-1)\times(n-1)}(\mathbb{R})$ obtained from A by deleting the ith row and the jth column. Then we can compute the determinant of A by *cofactor expansion*:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\widetilde{A}_{ij}) \quad \text{for } \textit{any } 1 \leq i \leq n,$$

i.e., we can use cofactor expansion along any row, not just the top row $i = 1$. Similarly, we can use cofactor expansion along any column to compute the determinant of A:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\widetilde{A}_{ij}) \quad \text{for } \textit{any } 1 \leq j \leq n.$$

Note that the cofactor expansion formulas reduce the computation of the determinant of an $n \times n$ matrix down to the computation of several $(n-1) \times (n-1)$ sized

determinants. Applying cofactor expansion recursively, eventually the computation will reduce to 2×2 or 1×1-sized determinants, which we know how to compute directly from above.

Example 5.4.4. Consider the 3×3 matrix

$$A = \begin{bmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{bmatrix} \in \mathrm{Mat}_{3\times 3}(\mathbb{R}).$$

We will calculate the determinant using cofactor expansion along the 1st row ($i = 1$):

$$\begin{aligned} \det \begin{bmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{bmatrix} &= (-1)^{1+1} A_{11} \det(\widetilde{A}_{11}) + (-1)^{1+2} A_{12} \det(\widetilde{A}_{12}) \\ &\quad + (-1)^{1+3} A_{13} \det(\widetilde{A}_{13}) \\ &= \det \begin{bmatrix} -5 & 2 \\ 4 & -6 \end{bmatrix} - 3 \det \begin{bmatrix} -3 & 2 \\ -4 & -6 \end{bmatrix} - 3 \det \begin{bmatrix} -3 & -5 \\ -4 & 4 \end{bmatrix} \\ &= \big[(-5)(-6) - 2 \cdot 4\big] - 3\big[(-3)(-6) - 2(-4)\big] \\ &\quad - 3\big[(-3)4 - (-5)(-4)\big] \\ &= 22 - 3 \cdot 26 - 3(-32) = 40. \end{aligned}$$

In general, when using cofactor expansion to compute determinants, it helps to judiciously pick a row or a column that has many zeros, if there is one.

Properties of the determinant. The determinant gives us an answer to Question 5.4.3:

Determinant Property 5.4.5. *Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. Then the following are equivalent:*

(1) $\det(A) \neq 0$
(2) $\mathrm{null}(A) = \{0\}$.

In other words, to check if the nullspace has a nontrivial vector in it, just compute the determinant and check if it is $\neq 0$ or $= 0$. As it turns out, the Determinant Property 5.4.5 is really the only thing we need to know about determinants going forward.

Nevertheless, here are some other properties of the determinant which might be useful for computing determinants:

Fact 5.4.6. Suppose $A, B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $\alpha \in \mathbb{R}$. Then

(1) $\det(I_{n\times n}) = 1$
(2) $\det(\alpha A) = \alpha^n A$
(3) if B is obtained from A by either switching two rows or switching two columns (but not both), then $\det(B) = -\det(A)$.

5.5. Eigenvalues and eigenvectors

Recall that the identity matrix I has the property that $I\mathbf{v} = \mathbf{v}$ for any (appropriately-sized) vector $\mathbf{v}$. Another way to say this is that the identity matrix "scales the vector

by $\lambda = 1$", e.g.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Likewise, the matrix αI will scale a vector by $\lambda = \alpha$, e.g.

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Along similar lines, a diagonal matrix will scale certain vectors, but possibly with different scaling factors depending on the vector:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = 2 \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = 3 \cdot \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

For this reason, diagonal matrices are often very nice matrices to work with (in general the operation of scaling is computationally easier than the operation of matrix multiplication).

The concepts of *eigenvalue*, *eigenvector*, *eigenspace*, and *eigenbasis* will allow us to treat any square matrix almost as if it were a diagonal matrix.

Definition 5.5.1. Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ is a square matrix and $\lambda \in \mathbb{R}$. We say that λ is an **eigenvalue** for A if there exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

If λ is an eigenvalue of A, then we call a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ which satisfies $A\mathbf{v} = \lambda\mathbf{v}$ an **eigenvector** of A associated to λ.

The goal of this section is to answer the following question:

Question 5.5.2. *Given $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, how do we*

(1) find all eigenvalues λ of A,

and for each eigenvalue λ how do we

(2) find all eigenvectors $\mathbf{v}$ associated to λ?

The answer to Question 5.5.2(1) actually follows quite nicely from the Determinant Property 5.4.5. Indeed, suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, $\lambda \in \mathbb{R}$ and note that we have the following equivalences:

$$
\begin{aligned}
& \lambda \text{ is an eigenvalue of } A \\
\iff & \text{there exists nonzero } \mathbf{v} \in \mathbb{R}^n \text{ such that } A\mathbf{v} = \lambda\mathbf{v} \\
\iff & \text{there exists nonzero } \mathbf{v} \in \mathbb{R}^n \text{ such that } A\mathbf{v} - \lambda\mathbf{v} = 0 \\
\iff & \text{there exists nonzero } \mathbf{v} \in \mathbb{R}^n \text{ such that } A\mathbf{v} - \lambda I\mathbf{v} = 0 \\
\iff & \text{there exists nonzero } \mathbf{v} \in \mathbb{R}^n \text{ such that } (A - \lambda I)\mathbf{v} = 0 \\
\iff & \text{null}(A - \lambda I) \neq \{0\} \\
\iff & \det(A - \lambda I) = 0, \text{ by the Determinant Property 5.4.5.}
\end{aligned}
$$

Comparing the first and last part of the equivalence gives us an answer to part (1) of our question:

Eigenvalue Theorem 5.5.3. *Suppose $A \in \text{Mat}_{n\times n}(\mathbb{R})$ and $\lambda \in \mathbb{R}$. Then the following are equivalent:*

(1) λ is an eigenvalue of A,
(2) $\det(A - \lambda I) = 0$.

In other words, the eigenvalues of A are zeros of the "function" $\det(A - \lambda I)$. As it turns out, the expression $\det(A - \lambda I)$ is always a polynomial in the variable λ. This polynomial has a special name:

Definition 5.5.4. Suppose $A \in \text{Mat}_{n\times n}(\mathbb{R})$. The polynomial[2]

$$p(\lambda) \;:=\; (-1)^n \det(A - \lambda I) \;=\; \det(\lambda I - A)$$

is called the **characteristic polynomial** of A, and the equation

$$p(\lambda) \;=\; 0$$

is called the **characteristic equation**.

Thus the Eigenvalue Theorem 5.5.3 states that the eigenvalues of A are precisely the zeros of its characteristic polynomial.

Example 5.5.5. Find the eigenvalues for the following matrix:

$$A \;=\; \begin{bmatrix} 4 & 0 & -2 \\ 1 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Solution. We will first determine the characteristic polynomial of A. Note that

$$
\begin{aligned}
\det(A - \lambda I) \;&=\; (-1)^3 \det \begin{bmatrix} 4-\lambda & 0 & -2 \\ 1 & 1-\lambda & 2 \\ 0 & 0 & 2-\lambda \end{bmatrix} \\
&=\; -(4-\lambda) \det \begin{bmatrix} 1-\lambda & 2 \\ 0 & 2-\lambda \end{bmatrix} + 2 \begin{bmatrix} 1 & 1-\lambda \\ 0 & 0 \end{bmatrix} \\
&\qquad \text{(using cofactor expansion along the top row)} \\
&=\; -(4-\lambda)(1-\lambda)(2-\lambda) \\
&=\; (\lambda-4)(\lambda-1)(\lambda-2).
\end{aligned}
$$

[2]The factor $(-1)^n$ ensures that the polynomial is monic.

Thus we get three distinct eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 4$. □

Next we turn our attention to finding eigenvectors corresponding to a particular eigenvalue. Suppose λ is an eigenvalue of A. We already saw that an eigenvector $\mathbf{v}$ is a nonzero vector such that $(A - \lambda I)\mathbf{v} = 0$. Thus $\mathbf{v} \in \text{null}(A - \lambda I)$. In fact, *every* nonzero vector in $\text{null}(A - \lambda I)$ is an eigenvector associated to λ. This motivates the following definition:

Definition 5.5.6. Suppose $A \in \text{Mat}_{n\times n}(\mathbb{R})$ and λ is an eigenvalue of A. We define the **eigenspace** of λ to be

$$E_\lambda \;:=\; \text{null}(A - \lambda I),$$

i.e., the eigenspace E_λ is the set of all eigenvectors associated to λ together with the zero vector[3].

Since an eigenspace is a nullspace, we know how to find a basis for it:

Example 5.5.7. Find all eigenvectors of the matrix

$$A \;=\; \begin{bmatrix} 4 & 0 & -2 \\ 1 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Solution. In Example 5.5.5 we found three distinct eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 4$. For each of these eigenvalues, we need to compute a basis of its eigenspace.

($\lambda_1 = 1$) We will compute a basis of

$$\text{null}(A - I) \;=\; \text{null}\begin{bmatrix} 3 & 0 & -2 \\ 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that

$$\left[\begin{array}{ccc|c} 3 & 0 & -2 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We see that x_2 is a free variable and thus the general solution is:

$$\begin{aligned} x_1 &= 0 \\ x_2 &= x_2 \\ x_3 &= 0 \end{aligned}$$

Thus we can express the eigenspace E_1 as

$$E_1 \;=\; \text{null}(A - I) \;=\; \text{span}\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$$

($\lambda_2 = 2$) We compute a basis of $\text{null}(A - 2I)$:

$$\left[\begin{array}{ccc|c} 2 & 0 & -2 & 0 \\ 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

[3]The zero vector is always included in every eigenspace, although the zero vector is never considered an eigenvector.

We see that x_3 is a free variable and the general solution is

$$\begin{aligned} x_1 &= x_3 \\ x_2 &= 3x_3 \\ x_3 &= x_3 \end{aligned}$$

Thus we can express the eigenspace E_2 as

$$E_2 = \text{null}(A - 2I) = \text{span}\left(\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}\right)$$

($\lambda_3 = 4$) We compute a basis of $\text{null}(A - 4I)$:

$$\left[\begin{array}{ccc|c} 0 & 0 & -2 & 0 \\ 1 & -3 & 2 & 0 \\ 0 & 0 & -2 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{ccc|c} 1 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We see that x_2 is a free variable and the general solution is

$$\begin{aligned} x_1 &= 3x_2 \\ x_2 &= x_2 \\ x_3 &= 0 \end{aligned}$$

Thus we can express the eigenspace E_4 as

$$E_4 = \text{null}(A - 4I) = \text{span}\left(\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}\right) \qquad \square$$

We have one final definition:

Definition 5.5.8. Suppose $A \in \text{Mat}_{n\times n}(\mathbb{R})$ is a square matrix. An **eigenbasis** of A is a basis of $\mathbb{R}^n$ which is composed of eigenvectors of A. In other words, a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ is an eigenbasis of A if

(1) $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for some λ_i, for each $i = 1, \ldots, n$.
(2) $\mathbb{R}^n = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$
(3) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

Here is a fact about eigenbases which we are happy to assume:

Fact 5.5.9. Suppose $A \in \text{Mat}_{n\times n}(\mathbb{R})$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, for some $k \leq n$. If

(1) β_i is a basis of E_{λ_i} for each $i = 1, \ldots, k$ and
(2) $|\beta_1| + |\beta_2| + \cdots + |\beta_k| = n$,

then $\beta := \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an eigenbasis of A. In particular, if $k = n$, then $\beta = \beta_1 \cup \cdots \cup \beta_n$ is always an eigenbasis (i.e., condition (2) is automatically satisfied).

Example 5.5.10. Find an eigenbasis of

$$A = \begin{bmatrix} 4 & 0 & -2 \\ 1 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

SOLUTION. In example 5.5.7 we found that E_1 had basis

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

E_2 had basis

$$\left\{ \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right\}$$

and E_4 had basis

$$\left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Then by Fact 5.5.9 the following is an eigenbasis of A:

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \right\} \qquad \square$$

We conclude this section with a few remarks:

Remark 5.5.11. Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$.

(1) It is possible that an eigenbasis of A does not exist. This can only happen when $p(\lambda)$ has repeated roots. We will see what to do in this situation in the next chapter.
(2) It is possible that some of the eigenvalues of A are complex. In this case, the corresponding eigenvectors will have complex entries, but otherwise everything else is the same. We will see what complex eigenvalues/vectors means for us in the next chapter.
(3) If all the roots of $p(\lambda)$ are distinct and real, then there will be n distinct real eigenvalues and thus an eigenbasis will always exist.

CHAPTER 6

Systems of differential equations

Up until this point, we have only considered differential equations with *one* unknown function $y(t)$, e.g.,

$$\begin{aligned} y' &= f(t,y) \\ y'' + p(t)y' + q(t)y &= g(t) \end{aligned}$$

Unfortunately, in real world problems you generally have many unknown variables you are interested in and you are rarely ever lucky enough to have just a single unknown. Therefore, just like with linear equations, we have to consider now differential equations with multiple unknown functions which might be entangled with each other in various ways.

In this final chapter, we will study *systems of differential equations*, i.e., multiple equations which relate multiple unknown functions and their derivatives. For the sake of time, we will focus on a very special case: *homogeneous linear first-order systems with constant coefficients*.

6.1. Homogeneous linear systems with constant coefficients

Here is a typical example of the type of system we will consider:

Example 6.1.1. What are the solutions to the following system:

$$\begin{aligned} x_1' &= x_1 + 2x_2 \\ x_2' &= 2x_1 + x_2 \end{aligned}$$

Here, a solution is a pair of functions $x_1(t), x_2(t)$ such that when you plug both functions in, then both equations are satisfied. One can easily check that the pair $x_1 = e^{-t}, x_2 = -e^{-t}$ and the pair $x_1 = e^{3t}, x_2 = e^{3t}$ are both solutions to the system. In fact, we will see that the set of all solutions is precisely the set of all linear combinations of these two pairs.

Our first goal is to learn how to solve systems like the one in Example 6.1.1 above. This requires basically two things:

(1) Reinterpret these systems in terms of linear algebra (i.e., column vectors and matrices)
(2) Exploit as much of the Chapter 5 material as possible to make computations as straightforward as possible.

Definition 6.1.2. A **homogeneous linear system of differential equations** (with constant coefficients) is a set of differential equations of the following form:

$$(\dagger) \qquad \begin{aligned} x_1'(t) &= a_{11}x_1(t) + \cdots + a_{1n}x_n(t) \\ x_2'(t) &= a_{21}x_1(t) + \cdots + a_{2n}x_n(t) \\ &\vdots \\ x_n'(t) &= a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) \end{aligned}$$

where each $a_{ij} \in \mathbb{R}$ and $x_1(x), \ldots, x_n(t)$ are unknown functions. A **solution** to the (†) is a collection of n differentiable functions $x_1, x_2, \ldots, x_n : I \to \mathbb{R}$ (where $I \subseteq \mathbb{R}$ is an interval) such that plugging these functions in to (†) makes each equation true.

We will prefer to write systems in terms of matrices and vectors, so we can rewrite (†) above as:

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

or even as:

$$\mathbf{x}' = A\mathbf{x}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Note that with this notation, a **solution** is now a vector-valued function

$$\mathbf{x}(t) : \mathbb{R} \to \mathbb{R}^n.$$

Example 6.1.3. We can rewrite Example 6.1.1 now as:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}' = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

or as just

$$\mathbf{x}' = A\mathbf{x} \quad \text{where} \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

We were given two distinct solutions, which we can now write as:

$$\mathbf{x}_0(t) = \begin{bmatrix} e^{-t} \\ -e^{-t} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_1(t) = \begin{bmatrix} e^{3t} \\ e^{3t} \end{bmatrix}$$

To verify $\mathbf{x}_0(t)$ is a solution, first we can compute the lefthand side:

$$\mathbf{x}_0'(t) = \begin{bmatrix} e^{-t} \\ -e^{-t} \end{bmatrix}' = \begin{bmatrix} e^{-t} \\ -e^{-t} \end{bmatrix} = -\mathbf{x}_0(t)$$

Next we compute the righthand side:

$$A\mathbf{x}_0(t) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} e^{-t} \\ -e^{-t} \end{bmatrix} = \begin{bmatrix} e^{-t} - 2e^{-t} \\ 2e^{-t} - e^{-t} \end{bmatrix} = \begin{bmatrix} -e^{-t} \\ e^{-t} \end{bmatrix} = -\mathbf{x}_0(t)$$

Since the lefthand side equals the righthand side, we see that $\mathbf{x}_0(t)$ is indeed a solution. ($\mathbf{x}_1(t)$ can be verified in a similar way).

Before proceeding with how to find the solutions to systems like (†), we will first say a few general things about what the set of all solutions can look like. Our first result should come as no surprise:

Proposition 6.1.4. *Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $\mathbf{x}_1(t), \ldots, \mathbf{x}_k(t)$ are solutions to the system*

$$\mathbf{x}' \;=\; A\mathbf{x}.$$

Then for every $c_1, \ldots, c_k \in \mathbb{R}$ the linear combination $c_1\mathbf{x}_1(t) + \cdots + c_k\mathbf{x}_k(t)$ is also a solution.

Proof idea. This is because "taking the derivative" and "multiplying on the left by A" are both linear operations. □

Definition 6.1.5. Suppose $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_k(t) : I \to \mathbb{R}^n$ are vector-valued functions. We say that $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_k(t)$ are **linearly independent** if for every $c_1, \ldots, c_k \in \mathbb{R}$, if

$$c_1\mathbf{x}_1(t) + \cdots + c_k\mathbf{x}_k(t) \;=\; 0 \quad \text{for all } t \in I,$$

then $c_1 = c_2 = \cdots = c_k = 0$.

Otherwise, we say that $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_k(t)$ are **linearly dependent**.

Fortunately, the next fact says that checking linear (in)dependence of vector-valued functions boils down to checking whether certain $\mathbb{R}^n$ vectors are linearly (in)dependent:

Fact 6.1.6. Suppose $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_k(t)$ are solutions to $\mathbf{x}' = A\mathbf{x}$. If there is some fixed t_0 such that the column vectors $\mathbf{x}_1(t_0), \ldots, \mathbf{x}_k(t_0) \in \mathbb{R}^n$ are linearly dependent (respectively, linearly independent), then the functions $\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_k(t)$ are linearly dependent (resp., linearly independent).

We can now state (without proof) the general theorem describing the structure of the set of all solutions.

Theorem 6.1.7. *Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ are n linearly independent solutions to*

$$\mathbf{x}' \;=\; A\mathbf{x}.$$

Then $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ form a fundamental set of solutions, i.e., if $\mathbf{x}_0(t)$ is an arbitrary solution, then there are (necessarily unique) $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$\mathbf{x}_0(t) \;=\; c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_n\mathbf{x}_n(t) \quad \textit{for every } t.$$

Therefore, just as with homogeneous second-order linear differential equations and homogeneous matrix equations, the goal for linear systems $\mathbf{x}' = A\mathbf{x}$ is to find an appropriate number of linearly independent solutions. We now proceed with actually computing solutions to equations $\mathbf{x}' = A\mathbf{x}$. The primary idea is the following:

Proposition 6.1.8. *Suppose $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, λ is an eigenvalue of A, and $\mathbf{v}$ is an eigenvector associated to λ. Then*

$$\mathbf{x}(t) \;:=\; e^{\lambda t}\mathbf{v}$$

is a solution to the system $\mathbf{x}' = A\mathbf{x}$ and satisfies the initial condition $\mathbf{x}(0) = \mathbf{v}$.

PROOF. Let $\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$ be as in the statement of the proposition. Note that the lefthand side yields:

$$\mathbf{x}'(t) \;=\; \left(e^{\lambda t}\mathbf{v}\right)' \;=\; (e^{\lambda t})'\mathbf{v} \;=\; \lambda e^{\lambda t}\mathbf{v} \;=\; \lambda \mathbf{x}(t)$$

Whereas the righthand side yields:

$$A\mathbf{x}(t) \;=\; Ae^{\lambda t}\mathbf{v} \;=\; e^{\lambda t}A\mathbf{v} \;=\; e^{\lambda t}\lambda\mathbf{v} \;=\; \lambda\mathbf{x}(t). \qquad \square$$

Example 6.1.9. Find all solutions to the linear system:

$$\mathbf{x}' \;=\; A\mathbf{x} \quad \text{where} \quad A \;=\; \begin{bmatrix} -4 & 6 \\ -3 & 5 \end{bmatrix}.$$

SOLUTION. Proposition 6.1.8 suggests we should first look for eigenvalues and eigenvectors of A. First we obtain the characteristic polynomial:

$$\begin{aligned} p(\lambda) \;&=\; \det \begin{bmatrix} -4-\lambda & 6 \\ -3 & 5-\lambda \end{bmatrix} \\ &=\; (-4-\lambda)(5-\lambda) - 6(-3) \\ &=\; -20 + 4\lambda - 5\lambda + \lambda^2 + 18 \\ &=\; \lambda^2 - \lambda - 2 \\ &=\; (\lambda - 2)(\lambda + 1). \end{aligned}$$

Thus our eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = -1$. Now we will find the associated eigenvectors:

($\lambda_1 = 2$) We will find a basis for $\text{null}(A - 2I)$. Solving the associated homogeneous equation yields:

$$\left[\begin{array}{cc|c} -6 & 6 & 0 \\ -3 & 3 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Thus we found one eigenvector:

$$\mathbf{v}_1 \;=\; \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

($\lambda_2 = -1$). We will find a basis for $\text{null}(A + I)$. Solving the associated homogeneous equation yields:

$$\left[\begin{array}{cc|c} -3 & 6 & 0 \\ -3 & 6 & 0 \end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Thus we found one eigenvector

$$\mathbf{v}_2 \;=\; \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Proposition 6.1.8 tells us that the following two vector-valued functions are solutions to $\mathbf{x} = A\mathbf{x}$:

$$\begin{aligned} \mathbf{x}_1(t) \;&:=\; e^{2t}\mathbf{v}_1 \;=\; e^{2t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \mathbf{x}_2(t) \;&:=\; e^{-t}\mathbf{v}_2 \;=\; e^{-t}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \end{aligned}$$

Since $\mathbf{v}_1, \mathbf{v}_2$ are linearly independent, Fact 6.1.6 tells us that $\mathbf{x}_1(t), \mathbf{x}_2(t)$ are linearly independent vector-valued functions. Finally, Theorem 6.1.7 tells us that the general solution to $\mathbf{x}' = A\mathbf{x}$ is:

$$\mathbf{x}(t; C_1, C_2) \;=\; C_1 e^{2t}\begin{bmatrix}1\\1\end{bmatrix} + C_2 e^{-t}\begin{bmatrix}2\\1\end{bmatrix} \;=\; \begin{bmatrix}C_1 e^{2t} + 2C_2 e^{-t}\\ C_1 e^{2t} + C_2 e^{-t}\end{bmatrix} \qquad \square$$

6.2. Planar systems

In this section we will take a closer look at the 2×2 case. In this case, the characteristic polynomial is a quadratic polynomial, so there are three cases: distinct real roots case, complex conjugate roots case, and double real root case. Furthermore, the double real root case splits into two cases (because of the linear algebra): an easy case and an interesting case. We will say what to do in all four of these cases.

Distinct real roots case. The first case is when $p(\lambda)$ has two distinct real eigenvalues. We first give an example and then proceed with a general statement.

Example 6.2.1. Find the general solution to the following linear system:

$$\mathbf{x}' \;=\; A\mathbf{x} \quad \text{where} \quad A \;=\; \begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}$$

Solution. First we need to find the eigenvalues and associated eigenvectors of A. The characteristic polynomial is

$$p(\lambda) \;=\; \det\begin{bmatrix}-1-\lambda & 1\\ 1 & -1-\lambda\end{bmatrix} \;=\; (-1-\lambda)^2 - 1 \;=\; \lambda^2 + 2\lambda \;=\; (\lambda+2)(\lambda-0)$$

Thus the eigenvalues are $\lambda_1 = -2$ and $\lambda_2 = 0$. Now we find the associated eigenvectors.

($\lambda_1 = -2$) We need to find a basis for $\text{null}(A + 2I)$. Note that

$$\left[\begin{array}{cc|c}1 & 1 & 0\\ 1 & 1 & 0\end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c}1 & 1 & 0\\ 0 & 0 & 0\end{array}\right]$$

This yields the following eigenvector:

$$\mathbf{v}_1 \;=\; \begin{bmatrix}-1\\1\end{bmatrix}$$

($\lambda_2 = 0$) We need to find a basis for $\text{null}(A - 0I) = \text{null}(A)$. Note that

$$\left[\begin{array}{cc|c}-1 & 1 & 0\\ 1 & -1 & 0\end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c}1 & -1 & 0\\ 0 & 0 & 0\end{array}\right]$$

This yields the following eigenvector:

$$\mathbf{v}_2 \;=\; \begin{bmatrix}1\\1\end{bmatrix}$$

Next, Proposition 6.1.8 tells us that the following are both solutions to $\mathbf{x}' = A\mathbf{x}$:

$$\mathbf{x}_1(t) \;=\; e^{\lambda_1 t}\mathbf{v_1} \;=\; e^{-2t}\begin{bmatrix}-1\\1\end{bmatrix}$$

$$\mathbf{x}_2(t) \;=\; e^{\lambda_2 t}\mathbf{v_2} \;=\; \begin{bmatrix}1\\1\end{bmatrix}$$

Next, since the two column vectors $\mathbf{x}_1(0) = \mathbf{v}_1$ and $\mathbf{x}_2(0) = \mathbf{v}_2$ are linearly independent, by Fact 6.1.6 it follows that the two solutions $\mathbf{x}_1(t)$ and $\mathbf{x_2}(t)$ are linearly independent. Thus, by Theorem 6.1.7 we conclude the general solution is

$$\mathbf{x}(t; C_1, C_2) \;=\; C_1\mathbf{x}_1(t) + C_2\mathbf{x}_2(t) \;=\; C_1e^{-2t}\begin{bmatrix}-1\\1\end{bmatrix} + C_2\begin{bmatrix}1\\1\end{bmatrix} \;=\; \begin{bmatrix}-C_1e^{-2t} + C_2\\ C_1e^{-2t} + C_2\end{bmatrix}$$

□

The general case works exactly the same way:

Theorem 6.2.2 (Distinct real roots)**.** *Suppose $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ has two distinct real eigenvalues $\lambda_1 \neq \lambda_2 \in \mathbb{R}$. Furthermore, suppose $\mathbf{v}_1$ is an eigenvector associated with λ_1 and $\mathbf{v}_2$ is an eigenvector associated with λ_2. Then the general solution to $\mathbf{x}' = A\mathbf{x}$ is*

$$\mathbf{x}(t; C_1, C_2) \;=\; C_1e^{\lambda_1 t}\mathbf{v}_1 + C_2e^{\lambda_2 t}\mathbf{v}_2.$$

Complex conjugate roots case. The next case we will consider is when $p(\lambda)$ has a complex conjugate pair of complex (non-real) roots.

Example 6.2.3. Find the general solution to Find the general solution to the following linear system:

$$\mathbf{x}' \;=\; A\mathbf{x} \quad \text{where} \quad A \;=\; \begin{bmatrix}0 & 1\\ -2 & 2\end{bmatrix}$$

Solution. First we need to find the eigenvalues and associated eigenvectors of A. The characteristic polynomial is

$$p(\lambda) \;=\; \det\begin{bmatrix}-\lambda & 1\\ -2 & 2-\lambda\end{bmatrix} \;=\; -\lambda(2-\lambda) + 2 \;=\; \lambda^2 - 2\lambda + 2$$

and so the eigenvalues are

$$\lambda_1, \lambda_2 \;=\; \frac{2 \pm \sqrt{4-8}}{2} \;=\; 1 \pm i$$

so $\lambda_1 = 1 + i$ and $\lambda_2 = 1 - i = \overline{\lambda_1}$.

($\lambda_1 = 1 + i$) We need to find a basis for $\mathrm{null}(A - (1+i)I)$. Note that

$$\left[\begin{array}{cc|c}-1-i & 1 & 0\\ -2 & 1-i & 0\end{array}\right] \xrightarrow{\text{to RREF}} \left[\begin{array}{cc|c}1 & (-1+i)/2 & 0\\ 0 & 0 & 0\end{array}\right]$$

This yields the following eigenvector:

$$\mathbf{v}_1 \;=\; \begin{bmatrix}(1-i)/2\\ 1\end{bmatrix}$$

However, for convenience, we can scale $\mathbf{v}_1$ by $1 + i$ and instead use:

$$\mathbf{v}_1 \;=\; \begin{bmatrix}1\\ 1+i\end{bmatrix}$$

($\lambda_2 = 1 - i$) In this case, since $\lambda_2 = \overline{\lambda_1}$, $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$, and $\overline{A} = A$, taking complex conjugates yields:

$$\overline{A\mathbf{v}_1} \;=\; \overline{\lambda_1\mathbf{v}_1} \implies A\overline{\mathbf{v}_1} \;=\; \lambda_2\overline{\mathbf{v}_1}$$

This yields the following eigenvector associated to λ_2:

$$\mathbf{v}_2 \;=\; \overline{\mathbf{v}_1} \;=\; \begin{bmatrix}1\\ 1-i\end{bmatrix}$$

Next, Proposition 6.1.8 tells us that the following are both solutions to $\mathbf{x}' = A\mathbf{x}$:

$$\begin{aligned}\mathbf{z}_1(t) &:= e^{\lambda_1 t}\mathbf{v}_1 = e^{(1+i)t}\begin{bmatrix}1\\1+i\end{bmatrix}\\ \mathbf{z}_2(t) &:= e^{\lambda_2 t}\mathbf{v}_2 = e^{(1-i)t}\begin{bmatrix}1\\1-i\end{bmatrix}\end{aligned}$$

However, we are not done yet since $\mathbf{z}_1(t)$ and $\mathbf{z}_2(t) = \overline{\mathbf{z}_1(t)}$ are complex-valued solutions and we are ultimately looking for two linearly independent real-valued solutions. To find real-valued solutions, we can essentially do the same trick we used for Theorem 4.2.9, i.e., taking the real- and imaginary-parts of $\mathbf{z}_1(t)$. To justify this, recall from Proposition 6.1.4 that the set of all solutions to $\mathbf{x}' = A\mathbf{x}$ is closed under linear combinations. Thus

$$\begin{aligned}\mathbf{x}(t) &:= \frac{\mathbf{z}_1(t)+\mathbf{z}_2(t)}{2} = \operatorname{Re}\big(\mathbf{z}_1(t)\big)\\ \mathbf{y}(t) &:= \frac{\mathbf{z}_1(t)-\mathbf{z}_2(t)}{2i} = \operatorname{Im}\big(\mathbf{z}_1(t)\big)\end{aligned}$$

are also both solutions. Now we will use Euler's formula to get a better description of $\mathbf{x}(t)$ and $\mathbf{y}(t)$. Note that

$$\begin{aligned}\mathbf{z}_1(t) &= e^{(1+i)t}\begin{bmatrix}1\\1+i\end{bmatrix}\\ &= e^t(\cos t + i\sin t)\left(\begin{bmatrix}1\\1\end{bmatrix} + i\begin{bmatrix}0\\1\end{bmatrix}\right)\\ &= e^t\left(\cos t\begin{bmatrix}1\\1\end{bmatrix} - \sin t\begin{bmatrix}0\\1\end{bmatrix}\right) + ie^t\left(\cos t\begin{bmatrix}0\\1\end{bmatrix} + \sin t\begin{bmatrix}1\\1\end{bmatrix}\right)\\ &= e^t\begin{bmatrix}\cos t\\ \cos t - \sin t\end{bmatrix} + ie^t\begin{bmatrix}\sin t\\ \cos t + \sin t\end{bmatrix}\end{aligned}$$

Taking real and imaginary parts yields:

$$\begin{aligned}\mathbf{x}(t) &= e^t\begin{bmatrix}\cos t\\ \cos t - \sin t\end{bmatrix}\\ \mathbf{y}(t) &= e^t\begin{bmatrix}\sin t\\ \cos t + \sin t\end{bmatrix}\end{aligned}$$

Finally, since $\mathbf{x}(0) = \begin{bmatrix}1\\1\end{bmatrix}$ and $\mathbf{y}(0) = \begin{bmatrix}0\\1\end{bmatrix}$ are linearly independent, it follows that $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are linearly independent. Thus the general solution to $\mathbf{x}' = A\mathbf{x}$ is

$$\mathbf{w}(t; C_1, C_2) = C_1\mathbf{x}(t) + C_2\mathbf{y}(t) = C_1e^t\begin{bmatrix}\cos t\\ \cos t - \sin t\end{bmatrix} + C_2e^t\begin{bmatrix}\sin t\\ \cos t + \sin t\end{bmatrix} \quad \square$$

The general case works exactly the same way.

Theorem 6.2.4 (Complex conjugate roots)**.** *Suppose $A \in \operatorname{Mat}_{2\times 2}(\mathbb{R})$ has complex conjugate eigenvalues $\lambda, \overline{\lambda} \notin \mathbb{R}$, and $\mathbf{w}$ is an eigenvector associated to λ. Then $\overline{\mathbf{w}}$ is an eigenvector associated with $\overline{\lambda}$. Furthermore:*

(1) *(Complex version) The general solution to* $\mathbf{x}' = A\mathbf{x}$ *in terms of complex-valued functions is:*

$$\mathbf{x}(t; C_1, C_2) \;=\; C_1 e^{\lambda t}\mathbf{w} + C_2 e^{\overline{\lambda} t}\overline{\mathbf{w}}$$

(2) *(Real version) The general solution to* $\mathbf{x}' = A\mathbf{x}$ *is terms of real-valued functions is:*

$$\mathbf{x}(t; C_1, C_2) \;=\; C_1 e^{\alpha t}(\cos\beta t\ \mathbf{v}_1 - \sin\beta t\ \mathbf{v}_2) + C_2 e^{\alpha t}(\sin\beta t\ \mathbf{v}_1 + \cos\beta t\ \mathbf{v}_2)$$

where $\lambda = \alpha + i\beta$ *and* $\mathbf{w} = \mathbf{v}_1 + i\mathbf{v}_2$.

Double real root easy case. We now turn our attention to the case when $p(\lambda) = (\lambda - \lambda_1)^2$, i.e., when the characteristic polynomial has only one root of multiplicity two. First, we point out that exactly one of two things can happen:

(1) (Easy case) Either we can find two linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ associated to λ_0. An example of this case is

$$A \;=\; \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_0 \end{bmatrix}$$

which has linearly independent eigenvectors:

$$\mathbf{v}_1 \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_1 \;=\; \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(actually any two linearly independent vectors in $\mathbb{R}^2$ would work for this).

(2) (Interesting case) Or we can only find one linearly independent eigenvector $\mathbf{v}_1 \in \mathbb{R}^2$ associated to λ_0. An example of this is case is

$$A \;=\; \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}$$

which has only one linearly independent eigenvector:

$$\mathbf{v}_1 \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We will first look at the easy case. We will actually be able to completely solve the easy case, due to the following fact:

Fact 6.2.5. Suppose $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ has one real eigenvalue λ of multiplicity two. Furthermore, suppose we can find two linearly independent eigenvectors associated to A. Then

$$A \;=\; \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

and

$$\mathbf{e}_1 \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 \;=\; \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form an eigenbasis of A.

Proof. Suppose $\mathbf{v}_1, \mathbf{v}_1 \in \mathbb{R}^2$ are two linearly independent eigenvectors of A associated to λ. Then $\mathrm{null}(A - \lambda I) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) = \mathbb{R}^2$. In particular, we know the following two vectors are also linearly independent eigenvectors of A associated to λ:

$$\mathbf{e}_1 \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 \;=\; \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Now suppose $a, b, c, d \in \mathbb{R}$ are such that

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then the condition $A\mathbf{e}_1 = \lambda \mathbf{e}_1$ tells us that $a = \lambda, c = 0$, and the condition $A\mathbf{e}_2 = \lambda \mathbf{e}_2$ tells us that $b = 0, d = \lambda$. Thus

$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \qquad \square$$

This yields the following:

Theorem 6.2.6 (Double real root; easy case)**.** *Suppose $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ has only one eigenvalue $\lambda \in \mathbb{R}$ (of multiplicity two). Furthermore, suppose we can find two linearly independent eigenvectors of A associated to λ. Then the general solution to $\mathbf{x}' = A\mathbf{x}$ is*

$$\mathbf{x}(t; C_1, C_2) = C_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + C_2 e^{\lambda t} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} C_1 e^{\lambda t} \\ C_2 e^{\lambda t} \end{bmatrix}$$

Proof. Let

$$\mathbf{x}_1(t) := e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\mathbf{x}_2(t) := e^{\lambda t} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

By assumption, the eigenspace of λ is two-dimensional, so it must be all of $\mathbb{R}^2$, Thus the following two vectors are eigenvectors associated to λ:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Thus by Proposition 6.1.8 both $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are solutions to $\mathbf{x}' = A\mathbf{x}$. Furthermore, since $\mathbf{x}_1(0), \mathbf{x}_2(0)$ are linearly independent, it follows that $\mathbf{x}_1(t), \mathbf{x}_2(t)$ are also linearly independent. Thus by Theorem 6.1.7 it follows that the general solution is

$$\mathbf{x}(t; C_1, C_2) = C_1\mathbf{x}_1(t) + C_2\mathbf{x}_2(t) = C_1 e^{\lambda t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + C_2 e^{\lambda t} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} C_1 e^{\lambda t} \\ C_2 e^{\lambda t} \end{bmatrix} \qquad \square$$

Double real root interesting case. We now proceed with the interesting case. We investigate it by example.

Example 6.2.7. Find the general solution to $\mathbf{x}' = A\mathbf{x}$, where

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Solution. We begin by finding the eigenvalues and associated eigenvectors of A. The characteristic polynomial is

$$p(\lambda) = \det \begin{bmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{bmatrix} = (1-\lambda)(1-\lambda) = (\lambda - 1)^2.$$

Thus $\lambda_1 = 1$ is the only eigenvalue (of multiplicity two). Now we find all of the associated eigenvectors, i.e., we compute a basis for $\mathrm{null}(A - 1I)$. Note that

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

is already in RREF. Since there is one free variable, there is only one linearly independent eigenvector:

$$\mathbf{v}_1 \;=\; \begin{bmatrix}1\\0\end{bmatrix}$$

This tells us that

$$\mathbf{x}_1(t) \;=\; e^t\begin{bmatrix}1\\0\end{bmatrix}$$

is a solution to $\mathbf{x}' = A\mathbf{x}$. We are not done yet because we still need a second linearly independent solution. However, it appears that we are stuck since we don't have any more linearly independent eigenvectors of A (i.e., A fails to have an eigenbasis).

The solution is to *guess* that $\mathbf{x}' = A\mathbf{x}$ has a solution of the form:

$$\mathbf{x}(t) \;=\; e^{\lambda_1 t}(\mathbf{v}_2 + t\mathbf{v}_1)$$

for some vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. Supposing we have a solution of this form, lets see what this means for the vectors $\mathbf{v}_1, \mathbf{v}_2$. Note that

$$\mathbf{x}'(t) \;=\; \lambda_1 e^{\lambda_1 t}(\mathbf{v}_2 + t\mathbf{v}_1) + e^{\lambda_1 t}\mathbf{v}_1 \;=\; e^{\lambda_1 t}\big((\lambda_1\mathbf{v}_2 + \mathbf{v}_1) + \lambda_1 t\mathbf{v}_1\big)$$

whereas

$$A\mathbf{x}(t) \;=\; e^{\lambda_t 1}(A\mathbf{v}_2 + tA\mathbf{v}_1)$$

Equating these expressions and dividing by $e^{\lambda_1 t}$ (which is never zero) yields

$$(\lambda_1\mathbf{v}_2 + \mathbf{v}_1) + \lambda_1 t\mathbf{v}_1 \;=\; A\mathbf{v}_2 + tA\mathbf{v}_1$$

Since this needs to be true for all t, this yields:

$$\begin{aligned} A\mathbf{v}_2 &= \lambda_1\mathbf{v}_2 + \mathbf{v}_1\\ A\mathbf{v}_1 &= \lambda_1\mathbf{v}_1. \end{aligned}$$

In other words, $\mathbf{v}_1$ must be an eigenvector associated to λ_1, and $\mathbf{v}_2$ must be a solution to the equation

$$(A - \lambda_1 I)\mathbf{v}_2 \;=\; \mathbf{v}_1.$$

We have already found above that

$$\mathbf{v}_1 \;=\; \begin{bmatrix}1\\0\end{bmatrix}$$

works as an eigenvector. Now we will solve the equation

$$(A - \lambda_1 I)\mathbf{v}_2 \;=\; \mathbf{v}_1.$$

Setting up the augmented matrix yields:

$$\left[\begin{array}{cc|c} 0 & 1 & 1\\ 0 & 0 & 0 \end{array}\right]$$

which is already in RREF. We find that x_1 is a free variable, x_2 is a pivot variable, and the general solution is

$$\begin{aligned} x_1 &= x_1\\ x_2 &= 1 \end{aligned}$$

which in vector form is

$$\begin{bmatrix}x_1\\x_2\end{bmatrix} \;=\; x_2\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}$$

Thus

$$\mathbf{v}_2 \;=\; \begin{bmatrix}0\\1\end{bmatrix}$$

is a particular vector that works. This gives us the solutions:

$$\mathbf{x}_2 \;=\; e^{\lambda_1 t}(\mathbf{v}_2 + t\mathbf{v_1}) \;=\; e^t\left(\begin{bmatrix}0\\1\end{bmatrix} + t\begin{bmatrix}1\\0\end{bmatrix}\right) \;=\; \begin{bmatrix}te^t\\e^t\end{bmatrix}$$

We conclude the general solution is

$$\begin{aligned}\mathbf{x}(t;C_1,C_2) \;&=\; C_1\mathbf{x}_1(t) + C_2\mathbf{x}_2(t) \;=\; C_1e^t\begin{bmatrix}1\\0\end{bmatrix} + C_2e^t\left(\begin{bmatrix}0\\1\end{bmatrix} + t\begin{bmatrix}1\\0\end{bmatrix}\right)\\ &=\; \begin{bmatrix}C_1e^t + C_2te^t\\ C_2e^t\end{bmatrix}\end{aligned}$$ □

The general situation works exactly the same way:

Theorem 6.2.8. *Suppose $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ has only one eigenvalue $\lambda \in \mathbb{R}$ (of multiplicity two). Furthermore, suppose we can only find one linearly independent eigenvector $\mathbf{v}_1$ of A associated to λ. Then the general solution to $\mathbf{x}' = A\mathbf{x}$ is*

$$\mathbf{x}(t;C_1,C_2) \;=\; C_1e^{\lambda t}\mathbf{v}_1 + C_2e^{\lambda t}(\mathbf{v}_2 + t\mathbf{v}_1)$$

where $\mathbf{v}_2 \in \mathbb{R}^2$ is any particular solution to the matrix equation $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$.

6.3. Higher-order linear equations

This section is a sequel to Chapter 4, specifically Sections 4.1 and **??**. There we considered second-order linear equations:

$$y'' + p(t)y' + q(t)y \;=\; g(t),$$

and specifically homogeneous second-order linear equations with constant coefficients:

$$y'' + py' + qy \;=\; 0 \quad \text{with } p, q \in \mathbb{R}.$$

In this section[1] we will discuss homogeneous nth order linear equations with constant coefficients:

$$y^{(n)} + a_1y^{(n-1)} + \cdots + a_{n-1}y' + a_ny \;=\; 0 \quad \text{with } a_1, \ldots, a_n \in \mathbb{R}.$$

We will solve these equations in a three-step process:

(1) Convert the nth order linear system (in one unknown function) to an $n\times n$ linear system (with n unknown functions).
(2) Solve the nth order linear system.
(3) Convert the solution back in terms of a solution of the original linear differential equation.

We begin with a fairly representative example:

Example 6.3.1. Find the general solution to:

$$y^{(4)} - 13y'' + 36y \;=\; 0.$$

[1] In [**2**, §9.8] they consider more general linear equations of the form $y^{(n)} + a_1(t)y^{(n-1)} + \cdots + a_{n-1}y' + a_n(t)y = F(t)$ which might not have constant coefficients and might be inhomogeneous with a nonconstant forcing term. For us we will restrict our discussion to the homogeneous constant coefficient case.

SOLUTION. This is an equation with one unknown function. The first thing we do is convert this into an equation with four unknown functions by introducing three more auxiliary variables. Note that we will have to deal with four derivatives of $y(t)$, so to turn this into a first-order linear system, we define $x_2(t) := y'(t)$, $x_3(t) := y''(t) = x_2'(t)$, and $x_4(t) := y'''(t) = x_3'(t)$. Finally, to make the notation uniform, we also set $x_1(t) := y(t)$. This gives us the obvious conditions:

$$\begin{aligned} x_1'(t) &= x_2(t) \\ x_2'(t) &= x_3(t) \\ x_3'(t) &= x_4(t) \end{aligned}$$

What about $x_4'(t) = y^{(4)}(t)$? The original differential equation itself tells us how to relate this to the lower derivatives:

$$x_4'(t) = 13y''(t) - 36y(t) = 13x_3(t) - 36x_1(t)$$

Combining these four equations yields the system:

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t) \\ x_4'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -36 & 0 & 13 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix}$$

Of course, ultimately we are only interested in the first unknown function $x_1(t)$, but this is a quantity which we can read off as the first unknown function in a solution to the above system. Let's proceed to solve this system.

The first step is to compute the characteristic polynomial:

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I) \\ &= \begin{bmatrix} -\lambda & 1 & 0 & 0 \\ 0 & -\lambda & 1 & 0 \\ 0 & 0 & -\lambda & 1 \\ -36 & 0 & 13 & -\lambda \end{bmatrix} \\ &= 36 \det \begin{bmatrix} 1 & 0 & 0 \\ -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \end{bmatrix} - 13 \det \begin{bmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} - \lambda \det \begin{bmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 0 & 0 & -\lambda \end{bmatrix} \\ &\qquad \text{cofactor expansion along bottom row} \\ &= 36 - 13\lambda^2 + \lambda^4 \\ &= (\lambda - 2)(\lambda + 2)(\lambda - 3)(\lambda + 3) \end{aligned}$$

This gives us four eigenvalues, $\lambda_1 = 2$, $\lambda_2 = -2$, $\lambda_3 = 3$, $\lambda_4 = -4$. Next we compute the corresponding eigenvectors (calculation omitted):

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ -4 \\ 8 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 3 \\ 9 \\ 27 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} -1 \\ 3 \\ -9 \\ 27 \end{bmatrix}$$

In this case we have a real eigenbasis, so the general solution to $\mathbf{x}' = A\mathbf{x}$ is:

$$\mathbf{x}(t; C_1, C_2, C_3, C_4) = e^{2t}\begin{bmatrix}1\\2\\4\\8\end{bmatrix} + C_2e^{-2t}\begin{bmatrix}-1\\2\\-4\\8\end{bmatrix} + C_3e^{3t}\begin{bmatrix}1\\3\\9\\27\end{bmatrix} + C_4e^{-3t}\begin{bmatrix}-1\\3\\-9\\27\end{bmatrix}$$

$$= \begin{bmatrix} C_1e^{2t} - C_2e^{-2t} + C_3e^{3t} - C_4e^{-3t} \\ 2C_1e^{2t}2C_2e^{-2t} + 3C_3e^{3t} + 3C_4e^{-3t} \\ 4C_1e^{2t} - 4C_2e^{-2t} + 9C_3e^{3t} - 9C_4e^{-3t} \\ 8C_1e^{2t} + 8C_2e^{-2t} + 27C_3e^{3t} + 27C_4e^{-3t} \end{bmatrix}$$

In particular, the general solution to $y^{(4)} - 13y'' + 36y = 0$ is

$$y(t; C_1, C_2, C_3, C_4) = x_1(t) = C_1e^{2t} - C_2e^{-2t} + C_3e^{3t} - C_4e^{-3t}$$

which we might as well instead write as

$$y(t) = C_1e^{2t} + C_2e^{-2t} + C_3e^{3t} + C_4e^{-3t} \qquad \square$$

In general, suppose we have an nth order homogeneous linear differential equation with constant coefficients:

$$y^{(n)} + a_1y^{(n-1)} + \cdots + a_{n-1}y' + a_ny = 0 \quad \text{with } a_1, \ldots, a_n \in \mathbb{R}.$$

Then we can introduce $n-1$ additional unknown functions to stand for the higher derivatives of y: $x_1(t) := y(t)$, $x_2(t) := x_1'(t) = y'(t)$, $x_3(t) := x_2'(t) = y''(t), \ldots, x_n(t) := x_{n-1}'(t) = y^{(n-1)}(t)$. This gives the equations:

$$\begin{aligned} x_1'(t) &= x_2(t) \\ x_2'(t) &= x_3(t) \\ &\vdots \\ x_{n-1}'(t) &= x_n(t) \end{aligned}$$

Additionally, we can relate $x_n'(t) = y^{(n)}(t)$ to the lower derivatives using the original differential equation:

$$x_n'(t) = -a_nx_1(t) - a_{n-1}x_2(t) - \cdots - a_1x_n(t)$$

We then form the linear system:

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_{n-1}'(t) \\ x_n'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \vdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{n-1}(t) \\ x_n(t) \end{bmatrix}$$

This gives us an $n \times n$ linear system of the form $\mathbf{x}' = A\mathbf{x}$. The matrix A in this context is called the **companion matrix**. The following sums up what is true in general:

Theorem 6.3.2. *Consider the nth order homogeneous linear differential equation with constant coefficients:*

(A) $$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y \;=\; 0 \quad \text{with } a_1, \ldots, a_n \in \mathbb{R}.$$

and let

(B) $$\mathbf{x}' \;=\; A\mathbf{x}$$

be the associated linear system.

(1) The following are equivalent:
  (a) $y(t)$ is a solution to (A)
  (b) the vector-valued function

$$\mathbf{x}(t) \;=\; \begin{bmatrix} y(t) \\ y'(t) \\ y''(t) \\ \vdots \\ y^{(n-1)}(t) \end{bmatrix}$$

  is a solution to (B).

(2) Suppose $y_1(t), \ldots, y_n(t)$ are solutions to (A). The following are equivalent:
  (a) $y_1(t), \ldots, y_n(t)$ are linearly independent (as real-valued functions)
  (b) The following vector-valued functions are linearly independent:

$$\begin{bmatrix} y_1(t) \\ y_1'(t) \\ y_1''(t) \\ \vdots \\ y_1^{(n-1)}(t) \end{bmatrix}, \ldots, \begin{bmatrix} y_n(t) \\ y_n'(t) \\ y_n''(t) \\ \vdots \\ y_n^{(n-1)}(t) \end{bmatrix}$$

  (c) For some t_0 the following determinant is nonzero:

$$\det \begin{bmatrix} y_1(t_0) & y_2(t_0) & \cdots & y_n(t_0) \\ y_1'(t_0) & y_2'(t_0) & \cdots & y_n'(t_0) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(t_0) & y_2^{n-1}(t_0) & \cdots & y_n^{(n-1)}(t_0) \end{bmatrix} \neq 0$$

  (d) For every t the following determinant is nonzero:

$$\det \begin{bmatrix} y_1(t) & y_2(t) & \cdots & y_n(t) \\ y_1'(t) & y_2'(t) & \cdots & y_n'(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(t) & y_2^{n-1}(t) & \cdots & y_n^{(n-1)}(t) \end{bmatrix} \neq 0$$

This motivates the following definition:

Definition 6.3.3. Let $y_1, \ldots, y_n : I \to \mathbb{R}$ be real-valued functions ($I \subseteq \mathbb{R}$ is an interval). We define the **Wronskian** of $y_1, \ldots, y_n$ to be the function

$$W(t) \;=\; \det \begin{bmatrix} y_1(t) & y_2(t) & \cdots & y_n(t) \\ y_1'(t) & y_2'(t) & \cdots & y_n'(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(t) & y_2^{n-1}(t) & \cdots & y_n^{(n-1)}(t) \end{bmatrix}$$

We can now summarize everything in terms only of the original differential equation:

Proposition 6.3.4. *Suppose $y_1(t), \ldots, y_n(t)$ are solutions to*

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y \;=\; 0 \quad \textit{with } a_1, \ldots, a_n \in \mathbb{R}.$$

Then $y_1, \ldots, y_n$ are linearly independent iff $W(t) \neq 0$ iff $W(t_0) \neq 0$ for some fixed t_0. In this case, the general solution is

$$y(t) \;=\; C_1 y_1(t) + C_2 y_2(t) + \cdots + C_n y_n(t).$$

Of course, we are sweeping a few explanations under the rug. However, at this point you should believe that everything is properly justified using routine linear algebra arguments similar to those needed for homogeneous second-order linear equations, homogeneous matrix equations, and homogeneous linear systems.

APPENDIX A

Special functions

In this appendix we will include an overview of relevant properties of common elementary functions which arise in calculus and differential equations. In general we will work within the realm of real numbers, although everything we say has an appropriate extension to the bigger world of complex numbers. However, we might occasionally have to refer to complex numbers every now and then.

A.1. Polynomials

A **polynomial** (in the single variable X) is an expression of the form:

$$p(X) \;=\; a_n X^n + a_{n-1}X^{n-1} + \cdots + a_2 X^2 + a_1 X + a_0 \quad (\text{where each } a_i \in \mathbb{R})$$

If $a_n \neq 0$, then we call n the **degree** of $p(X)$, denoted $\deg p = n$. We may also choose to write a polynomial in **summation notation**:

$$p(X) \;=\; \sum_{k=0}^{n} a_k X^k$$

We naturally construe a polynomial as a function $p : \mathbb{R} \to \mathbb{R}$ by declaring for $\alpha \in \mathbb{R}$:

$$p(\alpha) \;:=\; a_n \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_2 \alpha^2 + a_1 \alpha + a_0$$

Recall that given two polynomial $p(X) = \sum_{k=0}^{n} a_k X^k$ and $q(X) = \sum_{k=0}^{n} b_k X^k$, we can form their **sum**:

$$(p+q)(X) \;:=\; \sum_{k=0}^{n} (a_k + b_k) X^k$$

and their **product**:

$$(p \cdot q)(X) \;:=\; \sum_{k} \left(\sum_{i+j=k} a_i b_j \right) X^k$$

where the above sum ranges over all possible indices.

Polynomials are perhaps the most well-behaved type of function which shows up in calculus. Indeed:

Fact A.1.1. Suppose

$$p(X) \;=\; a_n X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0 \;=\; \sum_{k=0}^{n} a_k X^k$$

is a polynomial of degree n. Then the following facts are true about $p(X)$ as a function $p : \mathbb{R} \to \mathbb{R}$:

(1) p is continuous on all of $\mathbb{R}$. In particular, for every $\alpha \in \mathbb{R}$:

$$\lim_{x \to \alpha} p(x) \;=\; p(\alpha)$$

(2) The limits at infinity are computed as follows:
   (a) if $n = 0$, then
$$\lim_{x\to\infty} p(x) \;=\; \lim_{x\to-\infty} p(x) \;=\; a_0$$
   (b) if $n \geq 1$ is even, then
$$\lim_{x\to\infty} p(x) \;=\; \lim_{x\to-\infty} p(x) \;=\; \begin{cases}\infty & \text{if } a_n > 0\\ -\infty & \text{if } a_n < 0\end{cases}$$
   (c) if $n \geq 1$ is odd, then
$$\lim_{x\to\infty} p(x) \;=\; \begin{cases}\infty & \text{if } a_n > 0\\ -\infty & \text{if } a_n < 0\end{cases} \quad \text{and} \quad \lim_{x\to-\infty} p(x) \;=\; \begin{cases}-\infty & \text{if } a_n < 0\\ \infty & \text{if } a_n > 0\end{cases}$$
(3) p is differentiable on all of $\mathbb{R}$ with derivative
$$\frac{dp}{dX}(X) \;=\; na_nX^{n-1} + (n-1)a_{n-1}X^{n-2} + \cdots + 2a_2X + a_1$$
$$= \sum_{k=1}^{n} ka_kX^{k-1} \;=\; \sum_{k=0}^{n-1}(k+1)a_{k+1}X^k$$
(4) Since the derivative of a polynomial is again a polynomial, p is infinitely differentiable on all of $\mathbb{R}$,
(5) Define the degree $n+1$ polynomial:
$$P(X) \;:=\; \frac{a_n}{n+1}X^{n+1} + \frac{a_{n-1}}{n}X^n + \cdots + \frac{a_1}{2}X^2 + a_0X$$
$$= \sum_{k=1}^{n+1} \frac{a_{k-1}}{k}X^k \;=\; \sum_{k=0}^{n} \frac{a_k}{k+1}X^{k+1}$$
Then:
   (a) $P(X)$ is an antiderivative of $p(X)$, i.e.,
$$\frac{d}{dx}P(X) \;=\; p(X),$$
   (b) the indefinite integral of $p(X)$ is
$$\int p(X)\,dX \;=\; P(X) + C,$$
   (c) the definite integral of $p(X)$ is
$$\int_a^b p(X)\,dX \;=\; P(b) - P(a),$$
   for every $a, b \in \mathbb{R}$.

The following is an important theoretical tool for studying polynomials:

Fundamental Theorem of (Complex) Algebra A.1.2. *Suppose $n \geq 1$. Then for every polynomial*
$$p(X) \;=\; a_nX^n + a_{n-1}X^n - 1 + \cdots + a_1X + a_0$$
of degree n, there exists complex numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that
$$p(X) \;=\; a_n(X-\alpha_1)(X-\alpha_2)\cdots(X-\alpha_n).$$

The numbers $\alpha_1, \ldots, \alpha_n$ in A.1.2 need not be distinct. One (very minor) drawback of A.1.2 is that some of the roots might be complex numbers which are not real numbers. Since we usually want to stick to working entirely with real numbers, the following variant will be useful for us:

Fundamental Theorem of (Real) Algebra A.1.3. *Suppose $n \geq 1$. Then for every polynomial*

$$p(X) \;=\; a_n X^n + a_{n-1} X^n - 1 + \cdots + a_1 X + a_0$$

of degree n, there exists $r, s \in \mathbb{N}$ with $r + 2s = n$, and real numbers

$$\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s, \gamma_1, \ldots, \gamma_s \in \mathbb{R}$$

such that:

(1) p can be factored into linear and quadratic factors

$$p(X) \;=\; a_n \underbrace{(X - \alpha_1) \cdots (X - \alpha_r)}_{\textit{linear factors}} \underbrace{(X^2 + \beta_1 X + \gamma_1) \cdots (X^2 + \beta_s X + \gamma_s)}_{\textit{quadratic factors}},$$

and

(2) for each $i = 1, \ldots, s$, we have $\beta_i^2 - 4\gamma_i < 0$, i.e., the quadratic factor $X^2 + \beta_i X + \gamma_i$ does not have real roots.

Theorem A.1.3 is an easy consequence of Theorem A.1.2 since complex roots of polynomials occur in conjugate pairs. Combining these conjugate pairs together is what give rise to the quadratic factors.

When dealing with quadratic polynomials with no real roots, the following trick is essential:

Completing the Square A.1.4. *Suppose $a, b, c \in \mathbb{R}$ are arbitrary such that $a \neq 0$. Then*

$$aX^2 + bX + c \;=\; a\left(X + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a} \;=\; a\left[\left(X + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right]$$

If the discriminant $b^2 - 4ac < 0$ is negative, then the constant $(4ac - b^2)/4a^2 > 0$ is positive.

A.2. Rational functions

A **rational function** (in the single variable X) is an expression of the form

$$r(X) \;=\; \frac{a_m X^m + a_{m-1} X^{m-1} + \cdots + a_1 X + a_0}{b_n X^n + b_{n-1} X^{n-1} + \cdots + b_1 X + b_0} \quad (\text{where } a_i, b_j \in \mathbb{R})$$

i.e., a rational function is a quotient

$$r(X) \;=\; \frac{p(X)}{q(X)}$$

of polynomials, where $p(X) = a_m X^m + \cdots + a_0$ and $q(X) = b_n X^n + \cdots + b_0$.

Recall that given two rational functions $r_0(X) = p_0(X)/q_0(X)$ and $r_1(X) = p_1(X)/q_1(X)$, we can form their **sum**:

$$(r_0 + r_1)(X) \;:=\; \frac{p_0(X)q_1(X) + p_1(X)q_0(X)}{q_0(X)q_1(X)}$$

and their **product**:

$$(r_0 \cdot r_1)(X) \;:=\; \frac{p_0(X)p_1(X)}{q_0(X)q_1(X)}$$

Just as with polynomials, we naturally construe a rational function as a real-valued function. Since the denominator of a fraction is never allowed to be zero, the domain of $r(X) = p(X)/q(X)$ is:

$$\text{domain}(r) \;:=\; \big\{\alpha \in \mathbb{R} : q(\alpha) \neq 0\big\} \;\subseteq\; \mathbb{R}$$

Then we define the function $r : \text{domain}(r) \to \mathbb{R}$ by declaring for $\alpha \in \mathbb{R}$:

$$r(\alpha) \;:=\; \frac{p(\alpha)}{q(\alpha)}$$

Warning A.2.1. In general the domain of a rational function might exclude so-called *removable singularities*. For example, consider the following two rational functions:

$$r_0(X) \;:=\; \frac{(X+1)(X+2)}{(X+1)(X+3)} \quad \text{and} \quad r_1(X) \;:=\; \frac{X+2}{X+3}$$

Then as real-valued functions, we have

$$\text{domain}(r_0) \;=\; \mathbb{R} \setminus \{-1, -3\} \quad \text{and} \quad \text{domain}(r_1) \;=\; \mathbb{R} \setminus \{-3\}$$

i.e., r_0 is defined everywhere except -1 whereas r_1 is defined everywhere except -3. However, for every $\alpha \in \mathbb{R} \setminus \{-1, -3\}$, we have $r_0(\alpha) = r_1(\alpha)$. In other words, r_0 and r_1 are essentially the same real-valued function except that r_1 is defined at one more point than r_0 is. In some sense, the fact that r_0 does not have -1 in its domain is an artificial obstacle. It is due to the factor $x + 1$ occurring in both the numerator and denominator. Since this has no effect on the value of the function (since it contributes multiplication by 1), we can just cancel these factors out and gain an extra point where the function is defined. In practice, when working with rational functions, you always want to make sure that the numerator and the denominator have no common factors so that you can work with the largest possible "true" domain of the rational function.

In the rest of this section, we will ignore the issue of removable singularities. After polynomials, rational functions are the second best-behaved family of functions which show up in calculus:

Fact A.2.2. Suppose

$$r(X) \;=\; \frac{p(X)}{q(X)}$$

is a rational function with domain $D := \text{domain}(r)$. Then the following facts are true about $r(X)$ as a function $r : D \to \mathbb{R}$:

(1) r is continuous on all of D. In particular, for every $\alpha \in D$:

$$\lim_{x \to \alpha} r(x) \;=\; r(\alpha)$$

(2) r is differentiable on all of D with derivative

$$\frac{dr}{dX}(X) \;=\; \frac{q(X)\frac{dp}{dX}(X) - p(X)\frac{dq}{dX}(X)}{\big(q(X)\big)^2}$$

which is also a rational function with domain D.

(3) It follows that $r(X)$ is infinitely differentiable on D.

Integration of rational functions is a little bit more complicated and requires so-called *partial fraction decomposition*. First, some terminology:

Definition A.2.3. Suppose $r(X) = p(X)/q(X)$ is a rational function. We say that $r(X)$ is a **proper** rational function if $\deg p < \deg q$. Otherwise, we say that $r(X)$ is an **improper** rational function.

We have two versions of partial fraction decomposition, depending on whether every factor of the denominator is linear or not:

Partial Fraction Decomposition A.2.4 (Complex Case)**.** *Suppose*

$$r(X) \;=\; \frac{p(X)}{q(X)}$$

is a proper rational function with $\deg q = n$*. Then:*

(1) *By Theorem A.1.2 there exists a nonzero real number* $a \in \mathbb{R}$*, distinct complex numbers* $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$*, and positive integers* $n_1, \ldots, n_r \in \mathbb{N}$ *such that*
   (a) $n_1 + \cdots + n_r = n$*, and*
   (b) $q(X) = a(X - \alpha_1)^{n_1} \cdots (X - \alpha_r)^{n_r}$
(2) *there exists a family of complex numbers* $(A_{i,j})_{1 \le i \le r, 1 \le j \le n_i}$ *such that*

$$r(X) \;=\; \frac{p(X)}{q(X)} \;=\; \sum_{i=1}^{r} \sum_{j=1}^{n_i} \frac{A_{i,j}}{(X - \alpha_i)^j} \tag{A.1}$$

You should use A.2.4 any time every root of $q(X)$ is real, or if you want to work with complex numbers. If not every root of $q(X)$ is real and you want to avoid using complex numbers, then you should use the following:

Partial Fraction Decomposition A.2.5 (Real Case)**.** *Suppose*

$$r(X) \;=\; \frac{p(X)}{q(X)}$$

is a proper rational function with $\deg q = n$*. Then:*

(1) *By Theorem A.1.3 there exists* $r, s \in \mathbb{N}$ *such that* $r + 2s = n$*, a nonzero real numbers* $a \in \mathbb{R}$*, distinct real numbers* $\alpha_1, \ldots, \alpha_t \in \mathbb{R}$*, positive integers* $n_1, \ldots, n_t$*, distinct pairs of real numbers* $(\beta_1, \gamma_1), \ldots, (\beta_u, \gamma_u) \in \mathbb{R}^2$ *and positive integers* $n'_1, \ldots, n'_u$ *such that:*
   (a) $n_1 + \cdots + n_t = r$*,*
   (b) $n'_1 + \cdots + n'_u = s$*,*
   (c) *the denominator factors as:*

$$q(X) \;=\; a(X - \alpha_1)_1^n \cdots (X - \alpha_r)^{n_r} (X^2 + \beta_1 X + \gamma_1)^{n'_1} \cdots (X^2 + \beta_u X + \gamma_u)^{n'_u}$$

   (d) *for every* $i = 1, \ldots, u$*, we have* $\beta_i^2 - 4\gamma_i < 0$*, i.e., the quadratic factor* $X^2 + \beta_i X + \gamma_i$ *does not have real roots.*
(2) *There exists families of real numbers* $(A_{i,j})_{1 \le i \le r, 1 \le j \le n_i}$*,* $(B_{i,j})_{1 \le i \le s, 1 \le j \le n'_i}$*,* $(C_{i,j})_{1 \le i \le s, 1 \le j \le n'_i}$ *such that*

$$r(X) \;=\; \frac{p(X)}{q(X)} \;=\; \sum_{i=1}^{r} \sum_{j=1}^{n_i} \frac{A_{i,j}}{(X - \alpha_i)^j} + \sum_{i=1}^{s} \sum_{j=1}^{n'_i} \frac{B_{i,j} X + C_{i,j}}{(X^2 + \beta_i X + \gamma_i)^j} \tag{A.2}$$

For improper rational functions, we can write it as a polynomial plus a proper rational function:

Polynomial Division A.2.6. *Suppose $p(X)$ and $q(X)$ are polynomials:*

$$\begin{aligned} p(X) &= a_m X^m + \cdots + a_0 \\ q(X) &= b_n X^n + \cdots + b_0 \end{aligned}$$

with $\deg p = m \geq \deg q = n$, i.e., the rational function $r(X) = p(X)/q(X)$ is improper. Then:

(1) The following identity reduces the degree of the polynomial in the numerator:

$$\frac{p(X)}{q(X)} = \frac{a_m}{b_n} X^{m-n} + \frac{p(X) - (a_m/b_n)X^{m-n}q(X)}{q(X)}$$

where $\deg\big(p(X) - (a_m/b_n)X^{m-n}q(X)\big) < \deg p(X)$.

(2) By repeating (1) enough times, there are real numbers $c_{m-n}, c_{m-n-1}, \ldots, c_0 \in \mathbb{R}$ with $c_{m-n} \neq 0$, and a polynomial $\tilde{p}(X)$ with $\deg \tilde{p}(X) < n$, such that:

$$\frac{p(X)}{q(X)} = c_{m-n}X^{m-n} + c_{m-n-1}X^{m-n-1} + \cdots + c_1 X + c_0 + \frac{\tilde{p}(X)}{q(X)}$$

It follows that any rational function can be written as a polynomial (possibly zero) plus a partial fraction decomposition of the form (A.1) or (A.2). Once we decompose a rational function like this, then we can integrate it according to the following rules:

(1) Integrate the polynomial part according to Fact A.1.1(5).

(2) For functions of the form $1/(X-\alpha)$, $\alpha \in \mathbb{R}$, the indefinite integral is:

$$\int \frac{dX}{X-\alpha} = \ln|X-\alpha| + C$$

with domain $(-\infty, \alpha) \cup (\alpha, +\infty)$. Given $a < b \in \mathbb{R}$, the definite integral is::

$$\begin{cases} \int_a^b \frac{dX}{X-\alpha} = \ln(b-\alpha) - \ln(a-\alpha) & \text{if } \alpha < a < b \\ \int_a^b \frac{dX}{X-\alpha} = \ln(\alpha-b) - \ln(\alpha-a) & \text{if } a < b < \alpha \end{cases}$$

(3) For $n \geq 2$, functions of the form $1/(X-\alpha)^n$, $\alpha \in \mathbb{R}$, the indefinite integral is:

$$\int \frac{dX}{(X-\alpha)^n} = -\frac{1}{(n-1)(X-\alpha)^{n-1}} + C$$

with domain $(-\infty, \alpha) \cup (\alpha, +\infty)$. Given $a < b \in \mathbb{R}$ such that $\alpha < a < b$ or $a < b < \alpha$, the definite integral is:

$$\int_a^b \frac{dX}{(X-\alpha)^n} = \frac{1}{(n-1)(a-\alpha)^{n-1}} - \frac{1}{(n-1)(b-\alpha)^{n-1}}$$

(4) If $\beta, \gamma \in \mathbb{R}$ are such that $\beta^2 - 4\gamma < 0$, to compute the integral of $1/(X^2 + \beta X + \gamma)$, you first complete the square in the denominator:

$$\frac{1}{X^2+\beta X+\gamma} = \frac{1}{(X-\beta/2)^2 + (4\gamma-\beta^2)/4} = \frac{1}{(X-\beta/2)^2 + \delta}$$

(where $\delta := (4\gamma - \beta^2)/4$) and the integrate using arctangent. The indefinite integral is:

$$\begin{aligned}\int \frac{dX}{X^2+\beta X+\gamma} &= \int \frac{dX}{(X-\beta/2)^2+\delta} \\ &= \frac{1}{\sqrt{\delta}} \arctan\left(\frac{X-\beta/2}{\sqrt{\delta}}\right) + C\end{aligned}$$

with domain $\mathbb{R}$. Given $a < b \in \mathbb{R}$, the definite integral is:

$$\int_a^b \frac{dX}{X^2+\beta X+\gamma} = \frac{1}{\sqrt{\delta}}\left(\arctan\left(\frac{b-\beta/2}{\sqrt{\delta}}\right) - \arctan\left(\frac{a-\beta/2}{\sqrt{\delta}}\right)\right)$$

(5) If $\beta, \gamma \in \mathbb{R}$ are such that $\beta^2 - 4\gamma < 0$ and $B \in \mathbb{R}$, to compute the integral of $(X+B)/(X^2+\beta X+\gamma)$, you first complete the square in the denominator:

$$\frac{X+B}{X^2+\beta X+\gamma} = \frac{X+B}{(X-\beta/2)^2+\delta}$$

Then you rewrite the numerator into two parts:

$$\frac{X+B}{(X-\beta/2)^2+\delta} = \frac{1}{2}\frac{2(X-\beta/2)}{(X-\beta/2)^2+\delta} + \frac{B+\beta/2}{(X-\beta/2)^2+\delta}$$

The integral is the second part is done as in (4), the indefinite integral of the first part is:

$$\int \frac{1}{2}\frac{2(X-\beta/2)\,dX}{(X-\beta/2)^2+\delta} = \frac{1}{2}\ln\left|(X-\beta/2)^2+\delta\right| + C$$

with domain $\mathbb{R}$.

(6) If $\beta, \gamma \in \mathbb{R}$ are such that $\beta^2 - 4\gamma < 0$ and $n \geq 2$, to compute the integral of $1/(X^2+\beta X+\gamma)^n$, you first complete the square in the denominator:

$$\frac{1}{(X^2+\beta X+\gamma)^n} = \frac{1}{\left((X-\beta/2)^2+\delta\right)^n}$$

Then to compute the antiderivative, you first do the substitution $U = X - \beta/2$, $dU = dX$:

$$\int \frac{dX}{\left((X-\beta/2)^2+\delta\right)^n} = \int \frac{dU}{\left(U^2+\delta\right)^n}$$

Then you do the substitution $W = U/\sqrt{\delta}$, $dW = dU/\sqrt{\delta}$:

$$\int \frac{dU}{\left(U^2+\delta\right)^n} = \int \frac{\sqrt{\delta}\,dW}{\left((\sqrt{\delta}W)^2+\delta\right)^n} = \frac{\sqrt{\delta}}{\delta^n}\int \frac{dW}{(W^2+1)^n}$$

Then to compute $\int dW/(W^2+1)^n$ you use the trigonometric substitution $W = \tan\Theta$, $dW = \sec^2\Theta\,d\Theta$:

$$\begin{aligned}\int \frac{dW}{(W^2+1)^n} &= \int \frac{\sec^2\Theta\,d\Theta}{(\tan^2\Theta+1)^n} \\ &= \int \frac{\sec^2\Theta\,d\Theta}{\sec^{2n}\Theta} = \int \cos^{2n-2}\Theta\,d\Theta.\end{aligned}$$

At this point you use the rules for integrating powers of cosine.

(7) If $\beta, \gamma \in \mathbb{R}$ are such that $\beta^2 - 4\gamma < 0$, $B \in \mathbb{R}$, and $n \geq 2$, to compute the integral of $(X + B)/(X^2 + \beta X + \gamma)$ you complete the square and break up the numerator as in (5):

$$\frac{X + B}{\big((X - \beta/2)^2 + \delta\big)^n} \;=\; \frac{1}{2}\frac{2(X - \beta/2)}{\big((X - \beta/2)^2 + \delta\big)^n} + \frac{B + \beta/2}{\big((X - \beta/2)^2 + \delta\big)^n}$$

Then the second integral is computed as in (6), and the first integral is:

$$\int \frac{1}{2}\frac{2(X - \beta/2)\,dX}{\big((X - \beta/2)^2 + \delta\big)^n} \;=\; -\frac{1}{2(n-1)\big((X - \beta/2)^2 + \delta\big)^{n-1}}$$

A.3. Algebraic functions

A.4. The exponential function

The exponential function is the most important function in mathematics.[1] Here is its definition:

Definition A.4.1. Define the **exponential function** to be the function $\exp : \mathbb{R} \to \mathbb{R}$ defined by

$$\exp(\alpha) \;:=\; \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}$$

for every $\alpha \in \mathbb{R}$.

In general we will never use the definition of the exponential function explicitly in this class, we will only use known properties of the exponential function. Here are some basic properties of the exponential function:

Fact A.4.2. Suppose $\alpha, \beta \in \mathbb{R}$ are arbitrary. Then we have:

(1) $\exp(\alpha + \beta) = \exp(\alpha)\exp(\beta)$,
(2) $\exp(0) = 1$,
(3) exp is strictly increasing, i.e., if $\alpha < \beta$, then $\exp(\alpha) < \exp(\beta)$, and
(4) for every α, $\exp(\alpha) > 0$, and in particular, $\exp(\alpha) \neq 0$.

The exponential function is an extremely well-behave function in calculus:

Fact A.4.3. The function $\exp : \mathbb{R} \to \mathbb{R}$ has the following properties:

(1) exp is continuous. In particular, for every $\alpha \in \mathbb{R}$,

$$\lim_{x \to \alpha} \exp(x) \;=\; \exp(\alpha)$$

(2) the limits at $\pm\infty$ are as follows:

$$\lim_{x \to +\infty} \exp(x) \;=\; +\infty \quad \text{and} \quad \lim_{x \to -\infty} \exp(x) \;=\; 0.$$

(3) In particular, $\text{range}(\exp) = \{x \in \mathbb{R} : x > 0\} = (0, +\infty)$.
(4) exp is differentiable and

$$\frac{d}{dx}\exp(x) \;=\; \exp(x).$$

(5) It follows that exp is infinitely differentiable.

[1] See [**3**, pg. 1].

(6) The indefinite integral of exp is:

$$\int \exp(x)\,dx \;=\; \exp(x) + C$$

(7) Give $a < b \in \mathbb{R}$, the definite integral of exp is computed as:

$$\int_a^b \exp(x)\,dx \;=\; \exp(b) - \exp(a).$$

A.5. The logarithm

We saw in Section A.4 that the exponential function $\exp : \mathbb{R} \to (0, +\infty)$ is strictly increasing. In particular, it is invertible.

Definition A.5.1. We define the **logarithm** (or **natural logarithm**) to be the function $\ln : (0, +\infty) \to \mathbb{R}$ defined by:

$$\ln(y) \;=\; x \quad :\Longleftrightarrow \quad \exp(x) \;=\; y$$

for all $x \in \mathbb{R}$ and $y \in (0, +\infty)$. We also denote ln by log.

Here are some basic properties of the logarithm:

Fact A.5.2. Suppose $\alpha, \beta \in \mathbb{R}$ are arbitrary. Then we have:

(1) $\ln(\alpha\beta) = \ln\alpha + \ln\beta$,
(2) $\ln 1 = 0$, and
(3) ln is strictly increasing, i.e., if $\alpha < \beta$, then $\ln\alpha < \ln\beta$.

The logarithm is also a well-behaved function in calculus:

Fact A.5.3. The function $\ln : (0, +\infty) \to \mathbb{R}$ has the following properties:

(1) ln is continuous. In particular, for every $\alpha \in (0, +\infty)$,

$$\lim_{x \to \alpha} \ln x \;=\; \ln\alpha$$

(2) the limits at 0 and $+\infty$ are as follows:

$$\lim_{x \to 0+} \ln x \;=\; -\infty \quad \text{and} \quad \lim_{x \to +\infty} \ln x \;=\; +\infty.$$

(3) In particular, $\text{range}(\ln) = \mathbb{R}$.
(4) ln is differentiable and

$$\frac{d}{dx} \ln x \;=\; \frac{1}{x}$$

(5) It follows that ln is infinitely differentiable on $(0, +\infty)$.
(6) The indefinite integral of ln is:

$$\int \ln x\,dx \;=\; x\ln x - x + C,$$

where this family of antiderivatives is defined on $(0, +\infty)$.
(7) Given $0 < a < b \in \mathbb{R}$, the definite integral of ln is computed as:

$$\int_a^b \ln x\,dx \;=\; b\ln b - b - a\ln a + a$$

A.6. Power functions

A.7. Trigonometric functions

A.8. Inverse trigonometric functions

APPENDIX B

Foundations

Occasionally in this class we shall mention things like:

- Sets
- Operations on sets, like union, intersection,...
- Ordered pairs and cartesian products
- Relations and functions

For this class, you only need a working understanding of these concepts at the level of Math31B. However, we include a more rigorous treatment of these topics in this appendix if you desire a deeper understanding.

B.1. A Word about Definitions

When we write "$X := Y$", we mean that the object X does not have any meaning or definition yet, and we are defining X to be the same thing as Y. When we write "$X = Y$" we typically mean that the objects X and Y both already are defined and are the same. In other words, when writing "$X := Y$" we are performing an action (giving meaning to X) and when we write "$X = Y$" we are making an assertion of sameness.

In making definitions, we will often use the word "if" in the form "We say that ... if ..." or "If ..., then we say that ...". When the word "if" is used in this way in *definitions*, it has the meaning of "if and only if" (but only in definitions!). For example:

Definition B.1.1. Given integer d, n we say that d **divides** n if there exists an integer k such that $n = dk$.

This convention is followed in accordance with mathematical tradition. Also, we shall often write "iff" or "$\Leftrightarrow$" to abbreviate "if and only if." (Only mathematicians do this!)

B.2. Sets

A **set** is a collection of mathematical objects. Mathematical objects can be almost anything: numbers, other sets, functions, vectors, relations, matrices, graphs etc. For instance:

$$\{2, 5, 7\}, \quad \big\{3, 5, \{8, 9\}\big\}, \quad \text{and} \quad \{1, 3, 5, 7, \ldots\}$$

are all sets. A member of a set is called is called an **element** of the set. The membership relation is denoted with the symbol "$\in$", for instance, we write "$2 \in \{2, 5, 7\}$" (pronounced "2 is an element of the set $\{2, 5, 7\}$") to denote that the number 2 is a member of the set $\{2, 5, 7\}$. There are several ways to describe a set:

(1) by explicitly listing the elements in that set, i.e., the set $\{2, 5, 7\}$ is a set with three elements, the number 2, the number 5, and the number 7.

(2) by specifying a "membership requirement" that determines precisely which objects are in that set. For instance:

$$\big\{n \in \mathbb{Z} : \underbrace{n \text{ is positive and odd}}_{\text{membership requirement}}\big\}$$

is the set of all odd positive integers. The above set is pronounced "the set of all integers n such that n is positive and odd". The colon ":" is usually pronounced "such that", and the condition to the right of the colon is the membership requirement. Defining a set in this way is sometimes referred to as using **set-builder notation** since you are describing how the set is built (in the above example, the set is built by taking all integers and keeping the ones that are positive and odd), instead of explicitly specifying which elements are in the set. We could also choose to describe the set above by writing

$$\{1, 3, 5, 7, \ldots\},$$

although this might be a less ideal description because it requires the reader to guess or infer the meaning of "...".

The following is a very famous set:

Definition B.2.1. The **empty set** is the set which contains no elements (hence the name). It is denoted by either $\emptyset$ or $\{\}$.

The following are some of the main relationships two sets can have:

Definition B.2.2. Suppose A and B are sets. We say that

(1) A is a **subset** of B (notation: $A \subseteq B$) if every element of A is also an element of B, i.e.,
   - For every x, if $x \in A$, then $x \in B$

(2) A is **equal** to B (notation: $A = B$) if A and B have exactly the same elements, i.e.,
   - For every x, $x \in A$ if and only if $x \in B$

   equivalently, $A = B$ means the same thing as $A \subseteq B$ and $B \subseteq A$

(3) A is a **proper subset** of B (notation: $A \subsetneq B$) if $A \subseteq B$ and $A \neq B$.

Note that for any set A, we automatically have $\emptyset \subseteq A$.

Definition B.2.3. Given sets A and B, we define their **union** (notation: $A \cup B$) to be the set of all elements that are in either A or B, i.e.,

$$A \cup B \;:=\; \{x : x \in A \text{ or } x \in B\}.$$

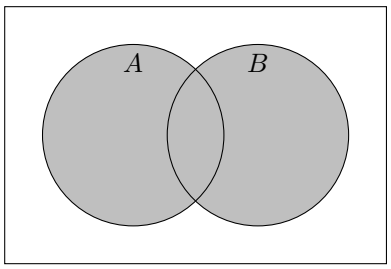


FIGURE B.1. Venn diagram of the union $A \cup B$ of the sets A and B

Definition B.2.4. Given sets A and B, we define their **intersection** (notation: $A \cap B$) to be the set of all elements they have in common, i.e.,

$$A \cap B \;:=\; \{x : x \in A \text{ and } x \in B\}.$$

We say that two sets A and B are **disjoint** if $A \cap B = \emptyset$.

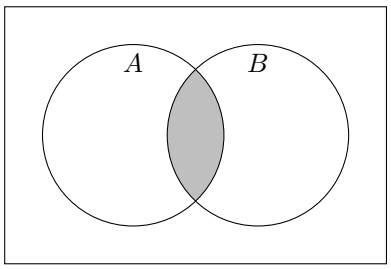


FIGURE B.2. Venn diagram of the intersection $A \cap B$ of the sets A and B

Definition B.2.5. Given sets A and B, we define their **(set) difference** (or **relative complement**) (notation: $A \setminus B$) to be the subset of A of all elements in A that are *not* in B, i.e.,

$$A \setminus B \;:=\; \{x : x \in A \text{ and } x \notin B\}.$$

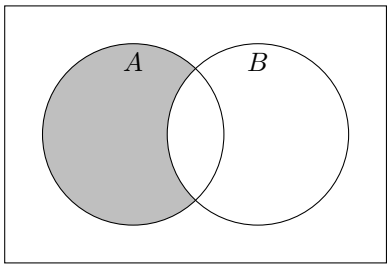


FIGURE B.3. Venn diagram of the difference $A \setminus B$ of the sets A and B

Suppose we have elements a, b, c, d such that $\{a, b\} = \{c, d\}$. It is tempting in this situation to conclude that "$a = c$ and $b = d$", but in general this is *false*. Indeed, we have $\{1, 2\} = \{2, 1\}$, but $1 \neq 2$ and $2 \neq 1$. This is because elements of a set are *unordered*. To get an *ordered* version of a two-element set we introduce the so-called *ordered pair* construction.

Definition B.2.6. Given objects a and b, we define their **ordered pair** to be the object:

$$(a, b) \;:=\; \big\{\{a\}, \{a, b\}\big\}$$

The righthand side of the definition might seem a little funny, but it guarantees the following:

Ordered Pair Property B.2.7. *For every a, b, c, d,*

$$(a, b) \;=\; (c, d) \quad \textit{if and only if} \quad a = c \textit{ and } b = d.$$

PROOF. Exercise! □

In practice, the Ordered Pair Property B.2.7 is really the only feature of ordered pairs that is ever relevant. You will almost never have to actually deal with the definition "$\{\{a\},\{a,b\}\}$", except when it comes proving the Ordered Pair Property.

Definition B.2.8. Given sets X and Y, we define the **cartesian product (of X and Y)** (notation: $X \times Y$) to be the following set:

$$X \times Y \;:=\; \big\{(x,y) : x \in X \text{ and } y \in Y\big\}$$

Example B.2.9. Suppose $X = \{0,1\}$ and $Y = \{a,b,c\}$. Then the cartesian product of X and Y is

$$X \times Y \;=\; \big\{(0,a),(0,b),(0,c),(1,a),(1,b),(1,c)\big\}.$$

Note that $|X| = 2$, $|Y| = 3$, and $|X \times Y| = 2 \cdot 3 = 6$.

The construction of pairs can be repeated:

Definition B.2.10. We define **ordered triples**, **ordered quadruples**, and more generally **ordered n-tuples** recursively as follows:

$$\begin{aligned}
(a_1,a_2,a_3) &:= \big((a_1,a_2),a_3\big)\\
(a_1,a_2,a_3,a_4) &:= \big((a_1,a_2,a_3),a_4\big)\\
&\vdots\\
(a_1,\dots,a_{n+1}) &:= \big((a_1,\dots,a_n),a_{n+1}\big)
\end{aligned}$$

for any objects $a_1, a_2, a_3, \dots$. It follows that two ordered n-tuples $(a_1,\dots,a_n)$ and $(b_1,\dots,b_n)$ are equal iff $a_i = b_i$ for each $i \in \{1,\dots,n\}$. Given sets $A_1,\dots,A_n$, we define their **n-fold cartesian product** to be the set

$$A_1 \times \cdots \times A_n \;:=\; \big\{(a_1,\dots,a_n) : a_i \in A_i \text{ for each } i = 1,\dots,n\big\}.$$

B.3. Relations

The mathematical structures we will deal with usually have more structure on it beyond the underlying set. For instance, we know that when we talk about the set $\mathbb{R}$, we also want to be able to talk about the linear order $\leq$ and the usual arithmetic binary functions $+$ and $\cdot$. If we didn't have these notions available to us, then there wouldn't be anything that special about the set $\mathbb{R}$ except that it's a very very large set. The formal way to make things like this is through *relations*.

Definition B.3.1. Given sets X and Y, we define a **(binary) relation on $X \times Y$** (or a **(binary) relation from X to Y**) to be a subset $R \subseteq X \times Y$. If R is a relation on $X \times Y$, then for an ordered pair $(x,y) \in X \times Y$ we will often write

$$\begin{aligned}
xRy &\quad \text{instead of } (x,y) \in R, \text{ and}\\
x\not{R}y &\quad \text{instead of } (x,y) \notin R.
\end{aligned}$$

(Note: xRy is pronounced "x is related to y (by R)"; and $x\not{R}y$ is pronounced "x is not related to y (by R)".)

Remark B.3.2. The word *binary* in Definition B.3.1 refers to the fact that R is a relation on a cartesian product on *two* sets: X and Y. One can also define *ternary relations* on $X \times Y \times Z$ and every *n-ary relations* on $X_1 \times X_2 \times \cdots \times X_n$. In this class we will (for the most part) restrict our attention to binary relations.

Example B.3.3. Consider $X := \{1, 2, 3, 4\}$ and $Y := \{a, b, c\}$ and the binary relation R on $X \times Y$ given by:

$$R \;=\; \big\{(1, a), (1, b), (2, a), (4, b), (4, c)\big\}$$

The relation R tells us, among other things, $1Ra$ but $3 \not R y$ for every $y \in R$. Since X, Y are small, we can picture all the relations specified by R with the following **arrow diagram**:

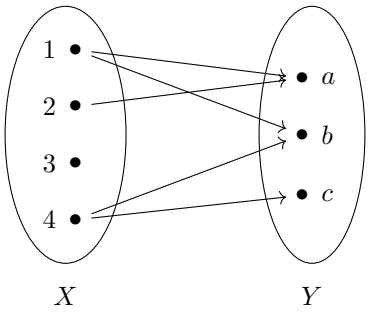


FIGURE B.4. Arrow diagram from X to Y illustrating the relation R on $X \times Y$

B.4. Functions

We are already familiar with functions $f\colon X \to Y$ as being some sort of machine that assigns to each input $x \in X$ a unique output $y \in Y$. The formal way to view functions is as a special case of relations:

Definition B.4.1. Suppose f is a relation on $X \times Y$. We say that f is a **function from X to Y** (notation: $f\colon X \to Y$) if for every $x \in X$ there is exactly one $y \in Y$ such that $(x, y) \in f$, i.e.,

(i) For each $x \in X$, there exists $y \in Y$ such that $(x, y) \in f$.
(ii) For each $x \in X$, and for every $y_1, y_2 \in Y$, if $(x, y_1) \in f$ and $(x, y_2) \in f$, then $y_1 = y_2$.

Note: (i) asserts there is *at least one* $y \in Y$, and (ii) asserts there is *at most one* $y \in Y$. Taken together, (i) and (ii) assert there is *exactly one* $y \in Y$ (with the property $(x, y) \in f$).

Suppose $f\colon X \to Y$. Then:

(1) We shall use the notation $f(x) = y$ to indicate that $(x, y) \in f$.
(2) The set X is called the **domain** of f (notation: $\mathrm{domain}(f) = X$).
(3) The set Y is called the **codomain** of f (notation: $\mathrm{codomain}(f) = Y$).
(4) The following subset of Y

$$\mathrm{range}(f) \;:=\; \big\{f(x) : x \in X\big\} \;=\; \big\{y \in Y : \text{there exists } x \in X \text{ such that } f(x) = y\big\}$$

is called the **range** of f.
(5) We also may use the notation "$x \mapsto f(x)\colon X \to Y$" instead of $f\colon X \to Y$, especially when the function f is determined by a formula in x and/or it is not necessary to give a name to the function; see Example B.4.2(2) below.

Example B.4.2.

(1) Given a set X we define the **identity function** on X (notation: $\mathrm{id}_X \colon X \to X$) to be the function that sends every $x \in X$ to itself, i.e.,

$$\mathrm{id}_X(x) \;:=\; x, \quad \text{for every } x \in X.$$

Note that in this case, $\mathrm{domain}(\mathrm{id}_X) = \mathrm{codomain}(\mathrm{id}_X) = \mathrm{range}(\mathrm{id}_X) = X$.

(2) The function

$$k \mapsto k^2 \colon \mathbb{Z} \to \mathbb{Z}$$

has domain $\mathbb{Z}$, codomain $\mathbb{Z}$ and range $\{0, 1, 4, 9, 16, \ldots\}$.

(3) The function

$$x \mapsto x^2 \colon \mathbb{R} \to \mathbb{R}$$

has domain $\mathbb{R}$, codomain $\mathbb{R}$ and range $\{y \in \mathbb{R} : y \geq 0\}$.

Question B.4.3. *What is the codomain of the following function:*

$$f \;:=\; \big\{(1, a), (2, c), (3, c), (4, b)\big\}$$

Answer B.4.4. Trick question! The domain is definitely the set $X := \{1, 2, 3, 4\}$, however, the *codomain* can technically be any set which contains $Y := \{a, b, c\}$. Indeed, f is a valid function of type "$X \to Y$" (in which case, the codomain would be Y), but it is also a valid function of type "$X \to Y \cup \{d, e, f\}$" (in which case, the codomain would be $Y \cup \{d, e, f\} = \{a, b, c, d, e, f\}$). The lesson here is that the codomain is determined by what we say it is when we are specifying the function as either $f \colon X \to Y$ or $f \colon X \to Y \cup \{d, e, f\}$. This annoyance only occurs for the *codomain*. The *domain* is always uniquely determined (as mentioned above) from the underlying set of ordered pairs, as is the *range* (which in this case is Y).

Just as with relations, we can form a new function from two given functions by *composition*.

Definition B.4.5. Suppose $f \colon X \to Y$ and $g \colon Y \to Z$ are functions. Then the **composition of g with f** is the function $g \circ f \colon X \to Z$ defined by:

$$(g \circ f)(x) \;:=\; g\big(f(x)\big) \;:=\; \text{the unique } z \in Z \text{ such that there is a } y \in Y \text{ such that } f(x) = y \text{ and } g(y) = z.$$

Remark B.4.6.

(1) Suppose we have three function $f \colon X \to Y$, $g \colon Y \to Z$ and $h : Z \to W$. Then we can create two new functions through composition: $g \circ f \colon X \to Z$ and $h \circ g \colon Y \to W$. Finally, we can create two new functions:

$$h \circ (g \circ f) : X \to W \quad \text{and} \quad (h \circ g) \circ f \colon X \to W.$$

It is a nice exercise to show that these functions are the same, i.e.,

$$h \circ (g \circ f) \;=\; (h \circ g) \circ f.$$

Thus we say that functional composition is *associative*.

(2) Functional composition allows us to highlight the two main properties of the identity function $\mathrm{id}_X : X \to X$:
   (a) For every function $f \colon X \to Y$ we have $f \circ \mathrm{id}_X = f$,
   (b) For every function $g : W \to X$ we have $\mathrm{id}_X \circ g = g$.

We can also (sometimes) consider the *inverse* of a function.

Definition B.4.7. Suppose $f\colon X \to Y$ is a function. We say that a function $g\colon Y \to X$ is an **inverse** to f if

$$f \circ g = \mathrm{id}_Y \quad \text{and} \quad g \circ f = \mathrm{id}_X .$$

We say that $f\colon X \to Y$ is an **invertible** function if there exists an inverse $g\colon Y \to X$.

At this point, it is not clear whether every function has an inverse (answer: no), or even in the cases when a function does have an inverse whether that inverse is unique (answer: yes). The following clears up the latter issue:

Lemma B.4.8 (Uniqueness of function inverse)**.** *Suppose $f\colon X \to Y$ is a function and $g, h\colon Y \to X$ are inverses to f. Then $g = h$.*

Proof. Note that

$$\begin{aligned} g &= g \circ \mathrm{id}_Y \quad \text{by Remark B.4.6(2)} \\ &= g \circ (f \circ h) \quad \text{since } h \text{ is an inverse of } f \\ &= (g \circ f) \circ h \quad \text{since composition is associative} \\ &= \mathrm{id}_X \circ h \quad \text{since } g \text{ is an inverse of } f \\ &= h \quad \text{by Remark B.4.6(2).} \end{aligned}$$

□

One special feature of the proof of Lemma B.4.8 is that it used very general principles (compositional property of identity, definition of inverse, associativity) and did not mention specific elements $x \in X$ at all. Analogues of this argument show up in many other areas of math, for example, in the proof that the inverse of an invertible matrix is unique. At any rate, we can now unambiguously define the inverse f^{-1} of an invertible function f:

Definition B.4.9. Suppose $f\colon X \to Y$ is an invertible function. Then we define $f^{-1}\colon Y \to X$ to be the (unique) inverse of f.

B.5. Three Special Types of Functions

There are three special flavors of functions which permeate all of mathematics:

Definition B.5.1. A function $f\colon X \to Y$ is called

(1) **injective** (or **one-to-one**) if for every $x_1, x_2 \in X$, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.
(2) **surjective** (tacitly: **surjective onto** Y) (or **onto**) if for every $y \in Y$ there exists an $x \in X$ such that $f(x) = y$. Equivalently, f is surjective if $\mathrm{range}(f) = \mathrm{codomain}(f)$.
(3) **bijective** (or a **bijection**, or **one-to-one and onto**) if f is both injective and surjective

Note that the notion of *surjective* (as well as *bijective*) only makes sense when it is clear what the codomain is. If you change what the codomain is, the function might change whether it is surjective or not. For instance, in Question B.4.3, the function $f\colon X \to Y$ is surjective, but the function $f\colon X \to Y \cup \{d, e, f\}$ is not surjective, even though the two f's have the same underlying set!

We give some simple examples of functions which either have or do not have each of these properties:

Example B.5.2.

(1) Suppose $X = \{a, b, c\}$ and $Y = \{d, e, f\}$. Then the function $f\colon X \to Y$ specified in Figure B.5 is a bijection, i.e., it is both injective and surjective.

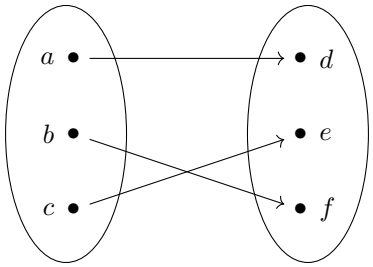


FIGURE B.5. A bijective (i.e., an injective and surjective) function

(2) Suppose $X = \{a, b, c\}$ and $Y = \{d, e\}$. Then the function $f\colon X \to Y$ specified in Figure B.6 is a surjective function but it is not injective.

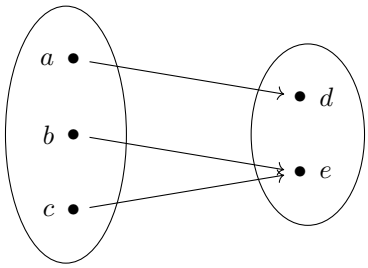


FIGURE B.6. A surjective function that is not bijective

(3) Suppose $X = \{a, b\}$ and $Y = \{c, d, e\}$. Then the function $f\colon X \to Y$ specified in Figure B.7 is an injective function but it is not surjective.

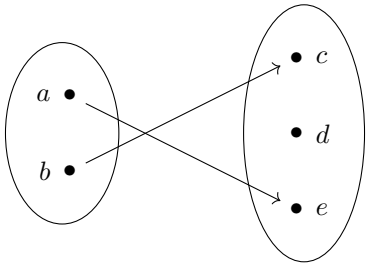


FIGURE B.7. An injective function that is not surjective

(4) Suppose $X = \{a, b\}$ and $Y = \{c, d\}$. Then the function $f\colon X \to Y$ specified in Figure B.8 is neither injective nor surjective.

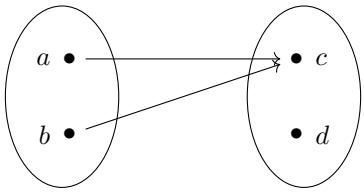


FIGURE B.8. A function that is neither injective nor surjective

These notions allow us to characterize which functions are invertible:

Theorem B.5.3. *Suppose $f\colon X \to Y$ is a function. The following are equivalent:*

(1) *f is a bijection.*
(2) *f is invertible.*

PROOF. Exercise! □

Bibliography

1. William A Adkins and Mark G Davidson, *First order differential equations*, Ordinary Differential Equations, Springer, 2012, pp. 1–100.
2. John C Polking, Albert Boggess, and David Arnold, *Differential equations with boundary value problems*, 2nd ed., Pearson/Prentice Hall, 2018.
3. Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987. MR 924157

Index