Notes on Diffy Qs

Differential Equations for MAT244

by Jiří Lebl

July 30, 2024
(version 6.7)

Typeset in LaTeX.

During the writing of this book, the author was in part supported by NSF grant DMS-0900885 and DMS-1362337.

The date is the main identifier of version. The major version / edition number is raised only if there have been substantial changes. Edition number started at 5, that is, version 5.0, as it was not kept track of before.

See https://www.jirka.org/diffyqs/ for more information (including contact information).

The LaTeX source for the book is available for possible modification and customization at github: https://github.com/jirilebl/diffyqs

Contents

Introduction **5**
- 0.1 Notes about these notes . . . 5
- 0.2 Introduction to differential equations . . . 7
- 0.3 Classification of differential equations . . . 14

1 First order equations **19**
- 1.1 Integrals as solutions . . . 19
- 1.2 Slope fields . . . 25
- 1.3 Separable equations . . . 31
- 1.4 Linear equations and the integrating factor . . . 38
- 1.5 Substitution . . . 44
- 1.6 Autonomous equations . . . 49
- 1.7 Numerical methods: Euler's method . . . 55
- 1.8 Exact equations . . . 61

2 Systems of ODEs **71**
- 2.1 Introduction to systems of ODEs . . . 71
- 2.2 Matrices and linear systems . . . 79
- 2.3 Linear systems of ODEs . . . 88
- 2.4 Eigenvalue method . . . 92
- 2.5 Two-dimensional systems and their vector fields . . . 99
- 2.6 Second order linear ODEs . . . 103
- 2.7 Second order systems and applications . . . 108
- 2.8 Multiple eigenvalues . . . 118

3 Nonlinear systems **125**
- 3.1 Linearization, critical points, and equilibria . . . 125
- 3.2 Stability and classification of isolated critical points . . . 131
- 3.3 Applications of nonlinear systems . . . 138
- 3.4 Limit cycles . . . 147
- 3.5 Chaos . . . 152

A Linear algebra **159**
A.1 Vectors, mappings, and matrices 159
A.2 Matrix algebra 170
A.3 Elimination 181
A.4 Subspaces, dimension, and the kernel 195
A.5 Inner product and projections 202
A.6 Determinant 210

B Complex numbers and Euler's formula **217**

Further Reading **219**

Solutions to Selected Exercises **221**

Index **229**

Introduction

0.1 Notes about these notes

Note: A section for the instructor.

This book originated from my class notes for Math 286 at the University of Illinois at Urbana-Champaign (UIUC) in Fall 2008 and Spring 2009. It is a first course on differential equations for engineers. Using this book, I also taught Math 285 at UIUC, Math 20D at University of California, San Diego (UCSD), and Math 2233 and 4233 at Oklahoma State University (OSU). Normally these courses are taught with Edwards and Penney, *Differential Equations and Boundary Value Problems: Computing and Modeling* [EP], or Boyce and DiPrima's *Elementary Differential Equations and Boundary Value Problems* [BD], and this book aims to be more or less a drop-in replacement. Other books I used as sources of information and inspiration are E.L. Ince's classic (and inexpensive) *Ordinary Differential Equations* [I], Stanley Farlow's *Differential Equations and Their Applications* [F], now available from Dover, Berg and McGregor's *Elementary Partial Differential Equations* [BM], and William Trench's free book *Elementary Differential Equations with Boundary Value Problems* [T]. See the Further Reading chapter at the end of the book.

0.1.1 Computer resources

The book's website `https://www.jirka.org/diffyqs/` contains the following resources:

1. Interactive SAGE demos.
2. Online WeBWorK homeworks (using either your own WeBWorK installation or Edfinity) for most sections, customized for this book.
3. The PDFs of the figures used in this book.

I taught the UIUC courses using IODE (`https://faculty.math.illinois.edu/iode/`). IODE is a free software package that works with Matlab (proprietary) or Octave (free software). The graphs in the book were made with the Genius software (see `https://www.jirka.org/genius.html`). I use Genius in class to show these (and other) graphs.

The LaTeX source of the book is also available for possible modification and customization at github (`https://github.com/jirilebl/diffyqs`).

0.1.2 Acknowledgments

Firstly, I would like to acknowledge Rick Laugesen. I used his handwritten class notes the first time I taught Math 286. My organization of this book through chapter 5, and the choice of material covered, is heavily influenced by his notes. Many examples and computations are taken from his notes. I am also heavily indebted to Rick for all the advice he has given me, not just on teaching Math 286. For spotting errors and other suggestions, I would also like to acknowledge (in no particular order): John P. D'Angelo, Sean Raleigh, Jessica Robinson, Michael Angelini, Leonardo Gomes, Jeff Winegar, Ian Simon, Thomas Wicklund, Eliot Brenner, Sean Robinson, Jannett Susberry, Dana Al-Quadi, Cesar Alvarez, Cem Bagdatlioglu, Nathan Wong, Alison Shive, Shawn White, Wing Yip Ho, Joanne Shin, Gladys Cruz, Jonathan Gomez, Janelle Louie, Navid Froutan, Grace Victorine, Paul Pearson, Jared Teague, Ziad Adwan, Martin Weilandt, Sönmez Şahutoğlu, Pete Peterson, Thomas Gresham, Prentiss Hyde, Jai Welch, Simon Tse, Andrew Browning, James Choi, Dusty Grundmeier, John Marriott, Jim Kruidenier, Barry Conrad, Wesley Snider, Colton Koop, Sarah Morse, Erik Boczko, Asif Shakeel, Chris Peterson, Nicholas Hu, Paul Seeburger, Jonathan McCormick, David Leep, William Meisel, Shishir Agrawal, Tom Wan, Andres Valloud, Martin Irungu, Justin Corvino, Tai-Peng Tsai, Santiago Mendoza Reyes, Glen Pugh, Michael Tran, Heber Farnsworth, Tamás Zsoldos, and probably others I have forgotten. Finally, I would like to acknowledge NSF grants DMS-0900885 and DMS-1362337.

0.2 Introduction to differential equations

Note: more than 1 lecture, §1.1 in [EP], chapter 1 in [BD]

0.2.1 Differential equations

The laws of physics are generally written down as differential equations. Therefore, all of science and engineering use differential equations to some degree. Understanding differential equations is essential to understanding almost anything you will study in your science and engineering classes. You can think of mathematics as the language of science, and differential equations are one of the most important parts of this language as far as science and engineering are concerned. As an analogy, suppose all your classes from now on were given in Swahili. It would be important to first learn Swahili, or you would have a very tough time getting a good grade in your classes.

You saw many differential equations already without perhaps knowing about it. And you even solved simple differential equations when you took calculus. Let us see an example you may not have seen:

$$\frac{dx}{dt} + x = 2\cos t. \tag{1}$$

Here x is the *dependent variable* and t is the *independent variable*. Equation (1) is a basic example of a *differential equation*. It is an example of a *first order differential equation*, since it involves only the first derivative of the dependent variable. This equation arises from Newton's law of cooling where the ambient temperature oscillates with time.

0.2.2 Solutions of differential equations

Solving the differential equation (1) means finding x in terms of t. That is, we want to find a function of t, which we call x, such that when we plug x, t, and $\frac{dx}{dt}$ into (1), the equation holds; that is, the left hand side equals the right hand side. It is the same idea as it would be for a normal (algebraic) equation of just x and t. We claim that

$$x = x(t) = \cos t + \sin t$$

is a *solution*. How do we check? We simply plug x into equation (1)! First we need to compute $\frac{dx}{dt}$. We find that $\frac{dx}{dt} = -\sin t + \cos t$. Now let us compute the left-hand side of (1).

$$\frac{dx}{dt} + x = \underbrace{(-\sin t + \cos t)}_{\frac{dx}{dt}} + \underbrace{(\cos t + \sin t)}_{x} = 2\cos t.$$

Yay! We got precisely the right-hand side. But there is more! We claim $x = \cos t + \sin t + e^{-t}$ is also a solution. Let us try,

$$\frac{dx}{dt} = -\sin t + \cos t - e^{-t}.$$

We plug into the left-hand side of (1)

$$\frac{dx}{dt} + x = \underbrace{(-\sin t + \cos t - e^{-t})}_{\frac{dx}{dt}} + \underbrace{(\cos t + \sin t + e^{-t})}_{x} = 2\cos t.$$

And it works yet again!

So there can be many different solutions. For this equation all solutions can be written in the form

$$x = \cos t + \sin t + Ce^{-t},$$

for some constant C. Different constants C will give different solutions, so there are really infinitely many possible solutions. See Figure 1 for the graph of a few of these solutions. We will see how we find these solutions a few lectures from now.

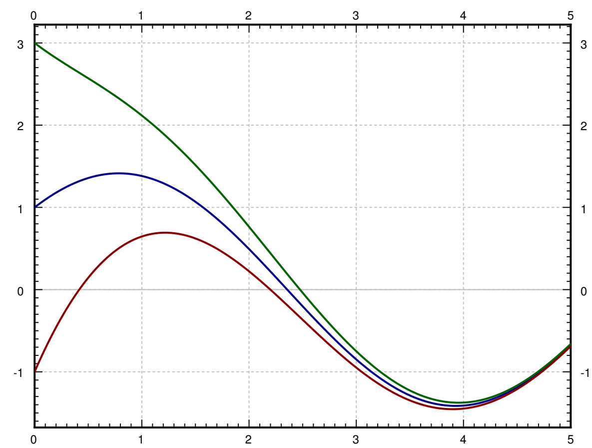


Figure 1: Few solutions of $\frac{dx}{dt} + x = 2\cos t$.

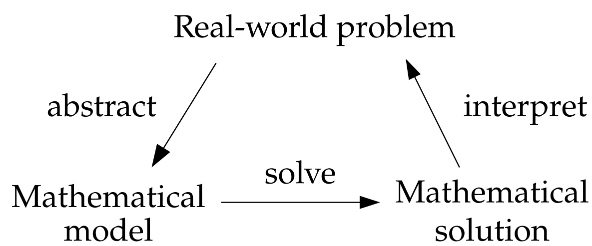
Solving differential equations can be quite hard. There is no general method that solves every differential equation. We will generally focus on how to get exact formulas for solutions of certain differential equations, but we will also spend a little bit of time on getting approximate solutions. And we will spend some time on understanding the equations without solving them.

Most of this book is dedicated to *ordinary differential equations* or ODEs, that is, equations with only one independent variable, where derivatives are only with respect to this one variable. If there are several independent variables, we get *partial differential equations* or PDEs.

Even for ODEs, which are very well understood, it is not a simple question of turning a crank to get answers. When you can find exact solutions, they are usually preferable to approximate solutions. It is important to understand how such solutions are found. Although in real applications you will leave much of the actual calculations to computers, you need to understand what they are doing. It is often necessary to simplify or transform your equations into something that a computer can understand and solve. You may even need to make certain assumptions and changes in your model to achieve this.

To be a successful engineer or scientist, you will be required to solve problems in your job that you never saw before. It is important to learn problem solving techniques, so that you may apply those techniques to new problems. A common mistake is to expect to learn some prescription for solving all the problems you will encounter in your later career. This course is no exception.

0.2.3 Differential equations in practice



So how do we use differential equations in science and engineering? First, we have some *real-world problem* we wish to understand. We make some simplifying assumptions and create a *mathematical model*. That is, we translate the real-world situation into a set of differential equations. Then we apply mathematics to get some sort of a *mathematical solution*. There is still something left to do. We have to interpret the results. We have to figure out what the mathematical solution says about the real-world problem we started with.

Learning how to formulate the mathematical model and how to interpret the results is what your physics and engineering classes do. In this course, we will focus mostly on the mathematical analysis. Sometimes we will work with simple real-world examples so that we have some intuition and motivation about what we are doing.

Let us look at an example of this process. One of the most basic differential equations is the standard *exponential growth model*. Let P denote the population of some bacteria on a Petri dish. We assume that there is enough food and enough space. Then the rate of growth of bacteria is proportional to the population—a large population grows quicker. Let t denote time (say in seconds) and P the population. Our model is

$$\frac{dP}{dt} = kP,$$

for some positive constant $k > 0$.

Example 0.2.1: Suppose there are 100 bacteria at time 0 and 200 bacteria 10 seconds later. How many bacteria will there be 1 minute from time 0 (in 60 seconds)?

First we need to solve the equation. We claim that a solution is given by

$$P(t) = Ce^{kt},$$

where C is a constant. Let us try:

$$\frac{dP}{dt} = Cke^{kt} = kP.$$

And it really is a solution.

OK, now what? We do not know C, and we do not know k. But we know something. We know $P(0) = 100$, and we know $P(10) = 200$. Let us plug these conditions in and see what happens.

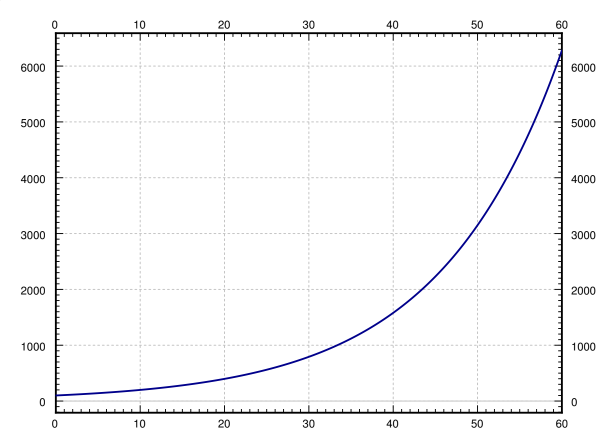


Figure 2: Bacteria growth in the first 60 seconds.

$$100 = P(0) = Ce^{k0} = C,$$
$$200 = P(10) = 100\,e^{k10}.$$

Therefore, $2 = e^{10k}$ or $k = \frac{\ln 2}{10} \approx 0.069$. So

$$P(t) = 100\, e^{(\ln 2)t/10} \approx 100\, e^{0.069t}.$$

At one minute, $t = 60$, the population is $P(60) = 6400$. See Figure 2 on the previous page.

Let us interpret the results. Does our solution mean that there must be exactly 6400 bacteria on the plate at 60s? No! We made assumptions that might not be true exactly, just approximately. If our assumptions are reasonable, then there will be approximately 6400 bacteria. Also, in real life P is a discrete quantity, not a real number. However, our model has no problem saying that for example at 61 seconds, $P(61) \approx 6859.35$.

Normally, the k in $P' = kP$ is given, and we want to solve the equation for different *initial conditions*. What does that mean? Take $k = 1$ for simplicity: We want to solve the equation $\frac{dP}{dt} = P$ subject to $P(0) = 1000$ (the initial condition). Then the solution is (exercise)

$$P(t) = 1000\, e^{t}.$$

We call $P(t) = Ce^{t}$ *the general solution*, as every solution of the equation can be written in this form for some constant C. We need an initial condition to find out what C is, in order to find the *particular solution* we are looking for. Generally, when we say "particular solution," we just mean some solution.

In real life, parameters such as k must first often be somehow computed or estimated. The example above shows how finding an analytic solution to the differential equation is useful in finding these parameters.

0.2.4 Four fundamental equations

A few equations appear often and it is useful to just memorize what their solutions are. Let us call them the four fundamental equations. Their solutions are reasonably easy to guess by recalling properties of exponentials, sines, and cosines. They are also simple to check, which is something that you should always do. No need to wonder if you remembered the solution correctly.

First such equation is

$$\frac{dy}{dx} = ky,$$

for some constant $k > 0$. Here y is the dependent and x the independent variable. The general solution for this equation is

$$y(x) = Ce^{kx}.$$

We saw above that this function is a solution, although we used different variable names.

Next,

$$\frac{dy}{dx} = -ky,$$

for some constant $k > 0$. The general solution for this equation is

$$y(x) = Ce^{-kx}.$$

Exercise* 0.2.1*: *Check that the y given is really a solution to the equation.*

Next, take the *second order differential equation*

$$\frac{d^2y}{dx^2} = -k^2 y,$$

for some constant $k > 0$. The general solution for this equation is

$$y(x) = C_1 \cos(kx) + C_2 \sin(kx).$$

Since the equation is a second order differential equation, we have two constants in our general solution.

Exercise* 0.2.2*: *Check that the y given is really a solution to the equation.*

Finally, consider the second order differential equation

$$\frac{d^2y}{dx^2} = k^2 y,$$

for some constant $k > 0$. The general solution for this equation is

$$y(x) = C_1 e^{kx} + C_2 e^{-kx},$$

or

$$y(x) = D_1 \cosh(kx) + D_2 \sinh(kx).$$

For those that do not know, cosh and sinh are defined by

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2}.$$

They are called the *hyperbolic cosine* and *hyperbolic sine*. These functions are sometimes easier to work with than exponentials. They have some nice familiar properties such as $\cosh 0 = 1$, $\sinh 0 = 0$, and $\frac{d}{dx} \cosh x = \sinh x$ (no that is not a typo) and $\frac{d}{dx} \sinh x = \cosh x$.

Exercise* 0.2.3*: *Check that both forms of the y given are really solutions to the equation.*

Example 0.2.2: In equations of higher order, you get more constants you must solve for to get a particular solution and hence you need more initial conditions. The equation $\frac{d^2y}{dx^2} = 0$ has the general solution $y = C_1 x + C_2$; simply integrate twice and don't forget about the constant of integration. Consider the initial conditions $y(0) = 2$ and $y'(0) = 3$. We plug in our general solution and solve for the constants:

$$2 = y(0) = C_1 \cdot 0 + C_2 = C_2, \qquad 3 = y'(0) = C_1.$$

In other words, $y = 3x + 2$ is the particular solution we seek.

An interesting note about cosh: The graph of cosh is the exact shape of a hanging chain. This shape is called a *catenary*. Contrary to popular belief this is not a parabola. If you invert the graph of cosh, it is also the ideal arch for supporting its weight. For example, the Gateway Arch in Saint Louis is an inverted graph of cosh—if it were just a parabola it might fall. The formula used in the design is inscribed inside the arch:

$$y = -127.7 \text{ ft} \cdot \cosh(x/127.7 \text{ ft}) + 757.7 \text{ ft}.$$

0.2.5 Exercises

Exercise* 0.2.4*: *Show that* $x = e^{4t}$ *is a solution to* $x''' - 12x'' + 48x' - 64x = 0$.

Exercise* 0.2.5*: *Show that* $x = e^{t}$ *is not a solution to* $x''' - 12x'' + 48x' - 64x = 0$.

Exercise* 0.2.6*: *Is* $y = \sin t$ *a solution to* $\left(\frac{dy}{dt}\right)^2 = 1 - y^2$*? Justify.*

Exercise* 0.2.7*: *Let* $y'' + 2y' - 8y = 0$*. Now try a solution of the form* $y = e^{rx}$ *for some (unknown) constant* r*. Is this a solution for some* r*? If so, find all such* r*.*

Exercise* 0.2.8*: *Verify that* $x = Ce^{-2t}$ *is a solution to* $x' = -2x$*. Find* C *to solve for the initial condition* $x(0) = 100$*.*

Exercise* 0.2.9*: *Verify that* $x = C_1e^{-t} + C_2e^{2t}$ *is a solution to* $x'' - x' - 2x = 0$*. Find* C_1 *and* C_2 *to solve for the initial conditions* $x(0) = 10$ *and* $x'(0) = 0$*.*

Exercise* 0.2.10*: *Find a solution to* $(x')^2 + x^2 = 4$ *using your knowledge of derivatives of functions that you know from basic calculus.*

Exercise* 0.2.11*: *Solve:*

a) $\dfrac{dA}{dt} = -10A, \quad A(0) = 5$

b) $\dfrac{dH}{dx} = 3H, \quad H(0) = 1$

c) $\dfrac{d^2y}{dx^2} = 4y, \quad y(0) = 0, \quad y'(0) = 1$

d) $\dfrac{d^2x}{dy^2} = -9x, \quad x(0) = 1, \quad x'(0) = 0$

Exercise* 0.2.12*: *Is there a solution to* $y' = y$*, such that* $y(0) = y(1)$*?*

Exercise* 0.2.13*: *The population of city X was 100 thousand 20 years ago, and the population of city X was 120 thousand 10 years ago. Assuming constant growth, you can use the exponential population model (like for the bacteria). What do you estimate the population is now?*

Exercise* 0.2.14*: *Suppose that a football coach gets a salary of one million dollars now, and a raise of* 10% *every year (so exponential model, like population of bacteria). Let* s *be the salary in millions of dollars, and* t *is time in years.*

a) What is $s(0)$ *and* $s(1)$*.*

b) Approximately how many years will it take for the salary to be 10 million.

c) Approximately how many years will it take for the salary to be 20 million.

d) Approximately how many years will it take for the salary to be 30 million.

Note: Exercises with numbers 101 and higher have solutions in the back of the book.

Exercise* 0.2.101*: *Show that* $x = e^{-2t}$ *is a solution to* $x'' + 4x' + 4x = 0$*.*

Exercise* 0.2.102*: *Is* $y = x^2$ *a solution to* $x^2y'' - 2y = 0$*? Justify.*

Exercise* 0.2.103*: *Let $xy'' - y' = 0$. Try a solution of the form $y = x^r$. Is this a solution for some r? If so, find all such r.*

Exercise* 0.2.104*: *Verify that $x = C_1 e^t + C_2$ is a solution to $x'' - x' = 0$. Find C_1 and C_2 so that x satisfies $x(0) = 10$ and $x'(0) = 100$.*

Exercise* 0.2.105*: *Solve $\frac{d\varphi}{ds} = 8\varphi$ and $\varphi(0) = -9$.*

Exercise* 0.2.106*: *Solve:*

a) $\dfrac{dx}{dt} = -4x, \quad x(0) = 9$

b) $\dfrac{d^2x}{dt^2} = -4x, \quad x(0) = 1, \quad x'(0) = 2$

c) $\dfrac{dp}{dq} = 3p, \quad p(0) = 4$

d) $\dfrac{d^2T}{dx^2} = 4T, \quad T(0) = 0, \quad T'(0) = 6$

0.3 Classification of differential equations

Note: less than 1 lecture or left as reading, §1.3 in [BD]

There are many types of differential equations, and we classify them into different categories based on their properties. Let us quickly go over the most basic classification. We already saw the distinction between ordinary and partial differential equations:

- *Ordinary differential equations* or (ODE) are equations where the derivatives are taken with respect to only one variable. That is, there is only one independent variable.
- *Partial differential equations* or (PDE) are equations that depend on partial derivatives of several variables. That is, there are several independent variables.

Let us see some examples of ordinary differential equations:

$$\frac{dy}{dt} = ky, \qquad \text{(Exponential growth)}$$
$$\frac{dy}{dt} = k(A-y), \qquad \text{(Newton's law of cooling)}$$
$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = f(t). \qquad \text{(Mechanical vibrations)}$$

And of partial differential equations:

$$\frac{\partial y}{\partial t} + c\frac{\partial y}{\partial x} = 0, \qquad \text{(Transport equation)}$$
$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad \text{(Heat equation)}$$
$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}. \qquad \text{(Wave equation in 2 dimensions)}$$

If there are several equations working together, we have a so-called *system of differential equations*. For example,

$$y' = x, \qquad x' = y$$

is a system of ordinary differential equations. Maxwell's equations for electromagnetics,

$$\nabla \cdot \vec{D} = \rho, \qquad \nabla \cdot \vec{B} = 0,$$
$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \qquad \nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t},$$

are a system of partial differential equations. The divergence operator $\nabla\cdot$ and the curl operator $\nabla\times$ can be written out in partial derivatives of the functions involved in the x, y, and z variables.

The next bit of information is the *order* of the equation (or system). The order is the order of the largest derivative that appears. If the highest derivative that appears is the first derivative, the equation is of first order. If the highest derivative that appears is the second derivative, then the equation is of second order. For example, Newton's law of cooling above is a first order equation, while the mechanical vibrations equation is a second order equation. The equation governing transversal vibrations in a beam,

$$a^4 \frac{\partial^4 y}{\partial x^4} + \frac{\partial^2 y}{\partial t^2} = 0,$$

is a fourth order partial differential equation. It is fourth order as at least one derivative is the fourth derivative. It does not matter that the derivative in t is only of second order.

In the first chapter, we will start attacking first order ordinary differential equations, that is, equations of the form $\frac{dy}{dx} = f(x, y)$. In general, lower order equations are easier to work with and have simpler behavior, which is why we start with them.

We also distinguish how the dependent variables appear in the equation (or system). In particular, an equation is *linear* if the dependent variable (or variables) and their derivatives appear linearly, that is, only as first powers, they are not multiplied together, and no other functions of the dependent variables appear. The equation is a sum of terms, where each term is a function of the independent variables or a function of the independent variables multiplied by a dependent variable or its derivative. Otherwise, the equation is called *nonlinear*. An ordinary differential equation is linear if it can be put into the form

$$a_n(x) \frac{d^n y}{dx^n} + a_{n-1}(x) \frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_1(x) \frac{dy}{dx} + a_0(x) y = b(x). \tag{2}$$

The functions $a_0, a_1, \ldots, a_n$ are called the *coefficients*. The equation is allowed to depend arbitrarily on the independent variable. So

$$e^x \frac{d^2 y}{dx^2} + \sin(x) \frac{dy}{dx} + x^2 y = \frac{1}{x} \tag{3}$$

is a linear equation as y and its derivatives only appear linearly.

All the equations and systems above as examples are linear. It may not be immediately obvious for Maxwell's equations unless you write out the divergence and curl in terms of partial derivatives. Let us see some nonlinear equations. For example Burger's equation,

$$\frac{\partial y}{\partial t} + y \frac{\partial y}{\partial x} = \nu \frac{\partial^2 y}{\partial x^2},$$

is a nonlinear second order partial differential equation. It is nonlinear because y and $\frac{\partial y}{\partial x}$ are multiplied together. The equation

$$\frac{dx}{dt} = x^2 \tag{4}$$

is a nonlinear first order differential equation as there is a second power of the dependent variable x. Another nonlinear ODE is the pendulum equation

$$\theta'' + \sin(\theta) = 0, \tag{5}$$

which is nonlinear as the dependent variable θ appears inside a sin function. Nonlinear equations are notoriously difficult to solve and their solutions may behave in strange an unexpected ways. Perhaps you have heard of chaos theory and the butterflies in the Amazon causing hurricanes in the Atlantic, all due to nonlinear equations. So sometimes we study related linear equations, such as $\theta'' + \theta = 0$ for the pendulum, instead.

A linear equation is further called *homogeneous* if all terms depend on the dependent variable. That is, if no term is a function of the independent variables alone. Otherwise, the equation is called *nonhomogeneous* or *inhomogeneous*. For example, the exponential growth equation, the wave equation, or the transport equation above are homogeneous. The mechanical vibrations equation above is nonhomogeneous as long as $f(t)$ is not the zero function. Similarly, if the ambient temperature A is nonzero, Newton's law of cooling is nonhomogeneous. A homogeneous linear ODE can be put into the form

$$a_n(x)\frac{d^n y}{dx^n} + a_{n-1}(x)\frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_1(x)\frac{dy}{dx} + a_0(x)y = 0.$$

Compare to (2) and notice there is no function $b(x)$.

If the coefficients of a linear equation are actually constant functions, then the equation is said to have *constant coefficients*. The coefficients are the functions multiplying the dependent variable(s) or one of its derivatives, not the function $b(x)$ standing alone. A constant coefficient nonhomogeneous ODE is an equation of the form

$$a_n\frac{d^n y}{dx^n} + a_{n-1}\frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_1\frac{dy}{dx} + a_0 y = b(x),$$

where $a_0, a_1, \ldots, a_n$ are all constants, but b may depend on the independent variable x. The mechanical vibrations equation above is a constant coefficient nonhomogeneous second order ODE. The same nomenclature applies to PDEs, so the transport equation, heat equation and wave equation are all examples of constant coefficient linear PDEs.

Finally, an equation (or system) is called *autonomous* if the equation does not depend on the independent variable. For autonomous ordinary differential equations, the independent variable is then thought of as time. Autonomous equation means an equation that does not change with time. For example, Newton's law of cooling is autonomous, so are the equations (4) and (5). On the other hand, mechanical vibrations (as long as $f(t)$ is nonconstant) or (3) are not autonomous. A general first order autonomous ODE would have the form

$$\frac{dx}{dt} = f(x).$$

0.3.1 Exercises

Exercise* 0.3.1*: *Classify the following equations. Are they ODE or PDE? Is it an equation or a system? What is the order? Is it linear or nonlinear, and if it is linear, is it homogeneous, constant coefficient? If it is an ODE, is it autonomous?*

a) $\sin(t)\dfrac{d^2x}{dt^2} + \cos(t)x = t^2$

b) $\dfrac{\partial u}{\partial x} + 3\dfrac{\partial u}{\partial y} = xy$

c) $y'' + 3y + 5x = 0, \quad x'' + x - y = 0$

d) $\dfrac{\partial^2 u}{\partial t^2} + u\dfrac{\partial^2 u}{\partial s^2} = 0$

e) $x'' + tx^2 = t$

f) $\dfrac{d^4x}{dt^4} = 0$

Exercise* 0.3.2*: *If $\vec{u} = (u_1, u_2, u_3)$ is a vector, we have the divergence $\nabla \cdot \vec{u} = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial z}$ and curl $\nabla \times \vec{u} = \left(\frac{\partial u_3}{\partial y} - \frac{\partial u_2}{\partial z}, \frac{\partial u_1}{\partial z} - \frac{\partial u_3}{\partial x}, \frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y}\right)$. Notice that curl of a vector is still a vector. Write out Maxwell's equations in terms of partial derivatives and classify the system.*

Exercise* 0.3.3*: *Suppose F is a linear function, that is, $F(x, y) = ax + by$ for constants a and b. What is the classification of equations of the form $F(y', y) = 0$.*

Exercise* 0.3.4*: *Write down an explicit example of a third order, linear, nonconstant coefficient, nonautonomous, nonhomogeneous system of two ODE such that every derivative that could appear, does appear.*

Exercise* 0.3.101*: *Classify the following equations. Are they ODE or PDE? Is it an equation or a system? What is the order? Is it linear or nonlinear, and if it is linear, is it homogeneous, constant coefficient? If it is an ODE, is it autonomous?*

a) $\dfrac{\partial^2 v}{\partial x^2} + 3\dfrac{\partial^2 v}{\partial y^2} = \sin(x)$

b) $\dfrac{dx}{dt} + \cos(t)x = t^2 + t + 1$

c) $\dfrac{d^7F}{dx^7} = 3F(x)$

d) $y'' + 8y' = 1$

e) $x'' + tyx' = 0, \quad y'' + txy = 0$

f) $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial s^2} + u^2$

Exercise* 0.3.102*: *Write down the general* zero*th order linear ordinary differential equation. Write down the general solution.*

Exercise* 0.3.103*: *For which k is $\frac{dx}{dt} + x^k = t^{k+2}$ linear. Hint: There are two answers.*

Chapter 1

First order equations

1.1 Integrals as solutions

Note: 1 lecture (or less), §1.2 in [EP], covered in §1.2 and §2.1 in [BD]

A first order ODE is an equation of the form

$$\frac{dy}{dx} = f(x, y),$$

or just

$$y' = f(x, y).$$

In general, there is no simple formula or procedure one can follow to find solutions. In the next few lectures, we will look at cases where solutions are not difficult to obtain. In this section, we consider the case when f is a function of x alone, that is, the equation is

$$y' = f(x). \tag{1.1}$$

We can integrate (antidifferentiate) both sides of the equation with respect to x:

$$\int y'(x)\,dx = \int f(x)\,dx + C,$$

that is,

$$y(x) = \int f(x)\,dx + C.$$

This $y(x)$ is actually the general solution. So to solve (1.1), we find some antiderivative of $f(x)$ and then we add an arbitrary constant to get the general solution.

Example 1.1.1: Find the general solution of $y' = 3x^2$.

Elementary calculus tells us that the general solution must be $y = x^3 + C$. Let us check by differentiating: $y' = 3x^2$. We got *precisely* our equation back.

Now is a good time to discuss a point about calculus notation and terminology. Calculus textbooks muddy the waters by talking about the integral as primarily the so-called indefinite integral. The indefinite integral is really the *antiderivative* (in fact the whole one-parameter family of antiderivatives). There really exists only one integral and that is the definite integral. The only reason for the indefinite integral notation is that we can always write an antiderivative as a (definite) integral. That is, by the fundamental theorem of calculus we can always write $\int f(x)\,dx + C$ as

$$\int_{x_0}^{x} f(t)\,dt + C.$$

Hence the terminology *to integrate* when we may really mean *to antidifferentiate*. Integration is just one way to compute the antiderivative (and it is a way that always works, see the following examples). Integration is defined as the area under the graph, it only happens to also compute antiderivatives. For sake of consistency, we will keep using the indefinite integral notation when we want an antiderivative, and you should *always* think of the definite integral as a way to write it.

Normally, we also have an initial condition such as $y(x_0) = y_0$ for some two numbers x_0 and y_0 (x_0 is often 0, but not always). We can then write the solution as a definite integral in a nice way. Suppose our problem is $y' = f(x)$, $y(x_0) = y_0$. Then the solution is

$$y(x) = \int_{x_0}^{x} f(t)\,dt + y_0. \tag{1.2}$$

Let us check! We compute $y' = f(x)$ via the fundamental theorem of calculus, and by Jupiter, y is a solution. Is it the one satisfying the initial condition? Well, $y(x_0) = \int_{x_0}^{x_0} f(t)\,dt + y_0 = y_0$. It is!

Do note that the definite integral and the indefinite integral (antidifferentiation) are completely different beasts. The definite integral always evaluates to a number. Therefore, (1.2) is a formula we can plug into the calculator or a computer, and it will be happy to calculate specific values for us. We will easily be able to plot the solution and work with it just like with any other function. It is not so crucial to always find a closed form for the antiderivative.

Example 1.1.2: Solve

$$y' = e^{-x^2}, \qquad y(0) = 1.$$

By the preceding discussion, the solution must be

$$y(x) = \int_{0}^{x} e^{-s^2}\,ds + 1.$$

Here is a good way to make fun of your friends taking second semester calculus. Tell them to find the closed form solution. Ha ha ha (bad math joke). It is not possible (in closed form). There is absolutely nothing wrong with writing the solution as a definite integral. This particular integral is in fact very important in statistics.

Using this method, we can also solve equations of the form

$$y' = f(y).$$

We write the equation in Leibniz notation:

$$\frac{dy}{dx} = f(y).$$

Now we use the inverse function theorem from calculus to switch the roles of x and y to obtain

$$\frac{dx}{dy} = \frac{1}{f(y)}.$$

What we are doing seems like algebra with dx and dy. It is tempting to just do algebra with dx and dy as if they were numbers. And in this case it does work. Be careful, however, as this sort of hand-waving calculation can lead to trouble, especially when more than one independent variable is involved. At this point, we can simply integrate with respect to y,

$$x(y) = \int \frac{1}{f(y)}\,dy + C.$$

Finally, we try to solve for y.

Example 1.1.3: Previously, we guessed $y' = ky$ (for some $k > 0$) has the solution $y = Ce^{kx}$. We could have found the solution by integrating. First we note that $y = 0$ is a solution. Henceforth, we assume $y \neq 0$. We write

$$\frac{dx}{dy} = \frac{1}{ky}.$$

We integrate in y to obtain

$$x(y) = x = \frac{1}{k}\ln|y| + D,$$

where D is an arbitrary constant. Now we solve for y (actually for $|y|$).

$$|y| = e^{kx-kD} = e^{-kD}e^{kx}.$$

If we replace e^{-kD} with an arbitrary constant C, we can get rid of the absolute value bars (which we can do as D was arbitrary). In this way, we also incorporate the solution $y = 0$. We get the same general solution as we guessed before, $y = Ce^{kx}$.

Example 1.1.4: Find the general solution of $y' = y^2$.

First we note that $y = 0$ is a solution. We can now assume that $y \neq 0$. Write

$$\frac{dx}{dy} = \frac{1}{y^2}.$$

We integrate to get

$$x = \frac{-1}{y} + C.$$

We solve for $y = \frac{1}{C-x}$. So the general solution is

$$y = \frac{1}{C-x} \qquad \text{together with} \qquad y = 0.$$

Note the singularities of the solution. If, for example, $C = 1$, then the solution "blows up" as we approach $x = 1$. See Figure 1.1. Generally, it is hard to tell from just looking at the equation itself how the solution is going to behave. The equation $y' = y^2$ is very nice and defined everywhere, but the solution is only defined on the interval $(-\infty, C)$ or (C, ∞). Usually when this happens, we only consider the solution on one of these intervals and not both. For example, if we impose an initial condition $y(0) = 1$, then the solution is $y = \frac{1}{1-x}$, and we would consider this solution only for x on the interval $(-\infty, 1)$. In the figure, it is the left side of the graph.

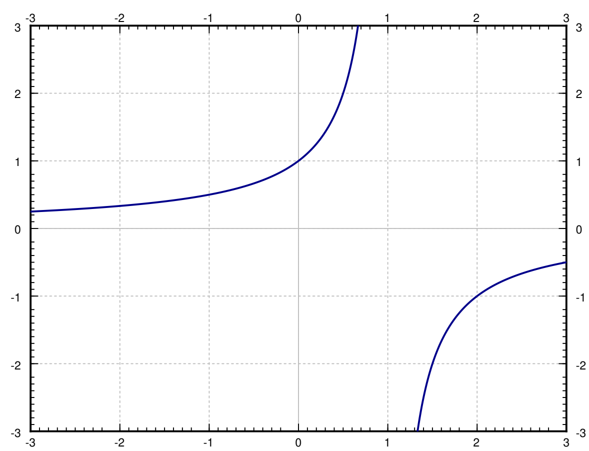


Figure 1.1: Plot of $y = \frac{1}{1-x}$.

Classical problems leading to differential equations solvable by integration are problems dealing with velocity, acceleration, and distance. You have surely seen these problems before in your calculus class.

Example 1.1.5: Suppose a car drives at a speed of $e^{t/2}$ meters per second, where t is time in seconds. How far did the car get in 2 seconds (starting at $t = 0$)? How far in 10 seconds?

Let x denote the distance the car traveled. The equation is

$$x' = e^{t/2}.$$

We just integrate this equation to get that

$$x(t) = 2e^{t/2} + C.$$

We still need to figure out C. We know that when $t = 0$, then $x = 0$. That is, $x(0) = 0$. So

$$0 = x(0) = 2e^{0/2} + C = 2 + C.$$

Thus $C = -2$ and

$$x(t) = 2e^{t/2} - 2.$$

Now we just plug in to get where the car is at 2 and at 10 seconds. We obtain

$$x(2) = 2e^{2/2} - 2 \approx 3.44 \text{ meters}, \qquad x(10) = 2e^{10/2} - 2 \approx 294 \text{ meters}.$$

Example 1.1.6: Suppose that the car accelerates at t^2 m/s². At time $t = 0$ the car is at the 1 meter mark and is traveling at 10 m/s. Where is the car at time $t = 10$?

Well this is actually a second order problem. If x is the distance traveled, then x' is the velocity, and x'' is the acceleration. The equation with initial conditions is

$$x'' = t^2, \qquad x(0) = 1, \qquad x'(0) = 10.$$

What if we say $x' = v$. Then we have the problem

$$v' = t^2, \qquad v(0) = 10.$$

Once we solve for v, we can integrate and find x.

***Exercise* 1.1.1:** *Solve for v, and then solve for x. Find $x(10)$ to answer the question.*

1.1.1 Exercises

***Exercise* 1.1.2:** *Solve $\frac{dy}{dx} = x^2 + x$ for $y(1) = 3$.*

***Exercise* 1.1.3:** *Solve $\frac{dy}{dx} = \sin(5x)$ for $y(0) = 2$.*

***Exercise* 1.1.4:** *Solve $\frac{dy}{dx} = \frac{1}{x^2-1}$ for $y(0) = 0$.*

***Exercise* 1.1.5:** *Solve $y' = y^3$ for $y(0) = 1$.*

***Exercise* 1.1.6** (little harder)**:** *Solve $y' = (y-1)(y+1)$ for $y(0) = 3$.*

***Exercise* 1.1.7:** *Solve $\frac{dy}{dx} = \frac{1}{y+1}$ for $y(0) = 0$.*

***Exercise* 1.1.8** (harder)**:** *Solve $y'' = \sin x$ for $y(0) = 0$, $y'(0) = 2$.*

***Exercise* 1.1.9:** *A spaceship is traveling at the speed $2t^2 + 1$ km/s (t is time in seconds). It is pointing directly away from earth and at time $t = 0$ it is 1000 kilometers from earth. How far from earth is it at one minute from time $t = 0$?*

***Exercise* 1.1.10:** *Solve $\frac{dx}{dt} = \sin(t^2) + t$, $x(0) = 20$. It is OK to leave your answer as a definite integral.*

Exercise* 1.1.11*: *A dropped ball accelerates downwards at a constant rate* 9.8 *meters per second squared. Set up the differential equation for the height above ground* h *in meters. Then supposing* $h(0) = 100$ *meters, how long does it take for the ball to hit the ground.*

Exercise* 1.1.12*: *Find the general solution of* $y' = e^x$*, and then* $y' = e^y$*.*

Exercise* 1.1.101*: *Solve* $\frac{dy}{dx} = e^x + x$ *and* $y(0) = 10$*.*

Exercise* 1.1.102*: *Solve* $x' = \frac{1}{x^2}$*,* $x(1) = 1$*.*

Exercise* 1.1.103*: *Solve* $x' = \frac{1}{\cos(x)}$*,* $x(0) = \frac{\pi}{4}$*.*

Exercise* 1.1.104*: *Sid is in a car traveling at speed* $10t + 70$ *miles per hour away from Las Vegas, where* t *is in hours. At* $t = 0$*, Sid is 10 miles away from Vegas. How far from Vegas is Sid 2 hours later?*

Exercise* 1.1.105*: *Solve* $y' = y^n$*,* $y(0) = 1$*, where* n *is a positive integer. Hint: You have to consider different cases.*

Exercise* 1.1.106*: *The rate of change of the volume of a snowball that is melting is proportional to the surface area of the snowball. Suppose the snowball is perfectly spherical. The volume (in centimeters cubed) of a ball of radius* r *centimeters is* $(4/3)\pi r^3$*. The surface area is* $4\pi r^2$*. Set up the differential equation for how the radius* r *is changing. Then, suppose that at time* $t = 0$ *minutes, the radius is 10 centimeters. After 5 minutes, the radius is 8 centimeters. At what time* t *will the snowball be completely melted?*

Exercise* 1.1.107*: *Find the general solution to* $y'''' = 0$*. How many distinct constants do you need?*

1.2 Slope fields

Note: 1 lecture, §1.3 in [EP], §1.1 in [BD]

As we said, the general first order equation we are studying looks like

$$y' = f(x, y).$$

A lot of the time, we cannot simply solve these kinds of equations explicitly. It would be nice if we could at least figure out the shape and behavior of the solutions, or find approximate solutions.

1.2.1 Slope fields

The equation $y' = f(x, y)$ gives you a slope at each point in the (x, y)-plane. And this is the slope a solution $y(x)$ would have at x if its value was y. In other words, $f(x, y)$ is the slope of a solution whose graph runs through the point (x, y). At a point (x, y), we draw a short line with the slope $f(x, y)$. For example, if $f(x, y) = xy$, then at point $(2, 1.5)$ we draw a short line of slope $xy = 2 \times 1.5 = 3$. If $y(x)$ is a solution and $y(2) = 1.5$, then the equation mandates that $y'(2) = 3$. See Figure 1.2.

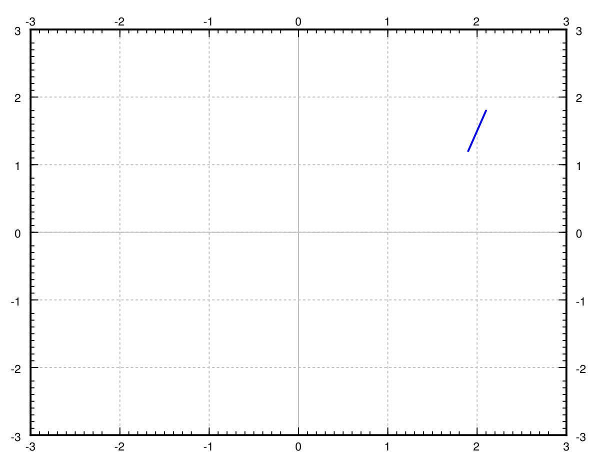


Figure 1.2: The slope $y' = xy$ at $(2, 1.5)$.

To get an idea of how solutions behave, we draw such lines at lots of points in the plane, not just the point $(2, 1.5)$. We would ideally want to see the slope at every point, but that is just not possible. Usually we pick a grid of points fine enough so that it shows the behavior, but not too fine so that we can still recognize the individual lines. We call this picture the *slope field* of the equation. See Figure 1.3 on the following page for the slope field of the equation $y' = xy$. In practice, one does not do this by hand, a computer can do the drawing.

Suppose we are given a specific initial condition $y(x_0) = y_0$. A solution, that is, the graph of the solution, would be a curve that follows the slopes we drew. For a few sample solutions, see Figure 1.4. It is easy to roughly sketch (or at least imagine) possible solutions in the slope field, just from looking at the slope field itself. You simply sketch a line that roughly fits the little line segments and goes through your initial condition.

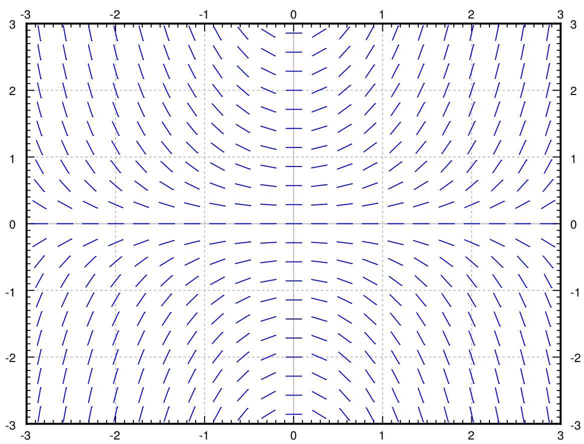


Figure 1.3: Slope field of $y' = xy$.

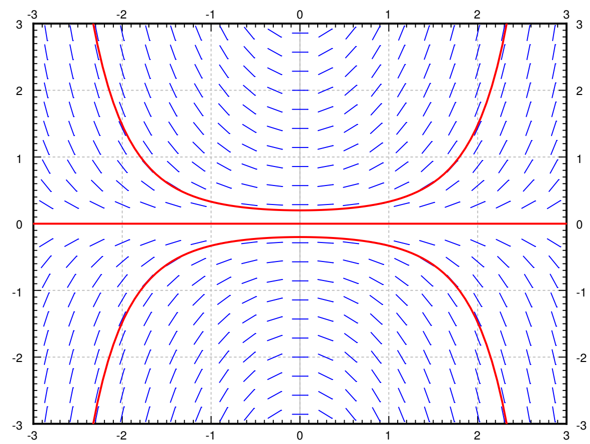


Figure 1.4: Slope field of $y' = xy$ with a graph of solutions satisfying $y(0) = 0.2$, $y(0) = 0$, and $y(0) = -0.2$.

By looking at the slope field we get a lot of information about the behavior of solutions without having to solve the equation. For example, in Figure 1.4 we see what the solutions do when the initial conditions are $y(0) > 0$, $y(0) = 0$ and $y(0) < 0$. A small change in the initial condition causes quite different behavior. We see this behavior just from the slope field and imagining what solutions ought to do.

We see a different behavior for the equation $y' = -y$. The slope field and a few solutions is in see Figure 1.5 on the next page. If we think of moving from left to right (perhaps x is time and time is usually increasing), then we see that no matter what $y(0)$ is, all solutions tend to zero as x tends to infinity. Again that behavior is clear from simply looking at the slope field itself.

1.2.2 Existence and uniqueness

We wish to ask two fundamental questions about the problem

$$y' = f(x, y), \qquad y(x_0) = y_0.$$

(i) Does a solution *exist*?

(ii) Is the solution *unique* (if it exists)?

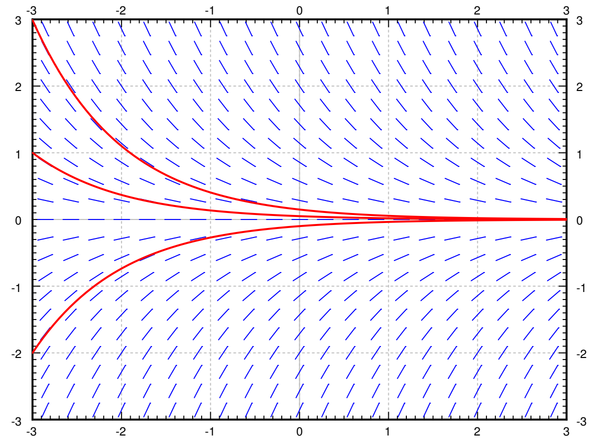


Figure 1.5: Slope field of $y' = -y$ with a graph of a few solutions.

What do you think is the answer? The answer seems to be yes to both does it not? Well, pretty much. But there are cases when the answer to either question can be no.

Since generally the equations we encounter in applications come from real life situations, it seems logical that a solution always exists. It also has to be unique if we believe our universe is deterministic. If the solution does not exist, or if it is not unique, we have probably not devised the correct model. Hence, it is good to know when things go wrong and why.

Example 1.2.1: Attempt to solve:

$$y' = \frac{1}{x}, \qquad y(0) = 0.$$

Integrate to find the general solution $y = \ln|x| + C$. The solution does not exist at $x = 0$. See Figure 1.6 on the following page. Moreover, the equation may have been written as the seemingly harmless $xy' = 1$.

Example 1.2.2: Solve:

$$y' = 2\sqrt{|y|}, \qquad y(0) = 0.$$

See Figure 1.7 on the next page. Note that $y = 0$ is a solution. But another solution is the function

$$y(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ -x^2 & \text{if } x < 0. \end{cases}$$

It is hard to tell by staring at the slope field that the solution is not unique. Is there any hope? Of course there is. We have the following theorem, known as Picard's theorem*.

*Named after the French mathematician Charles Émile Picard (1856–1941)

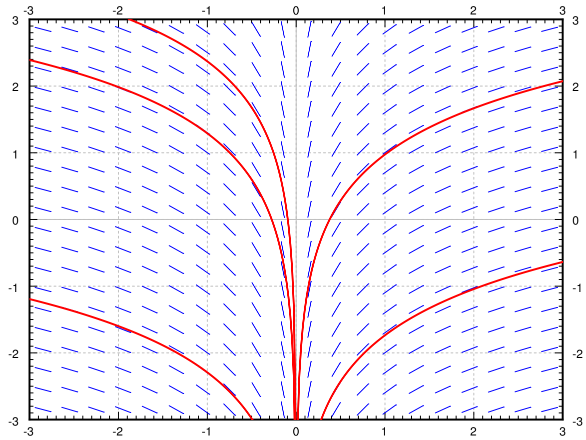


Figure 1.6: Slope field of $y' = 1/x$.

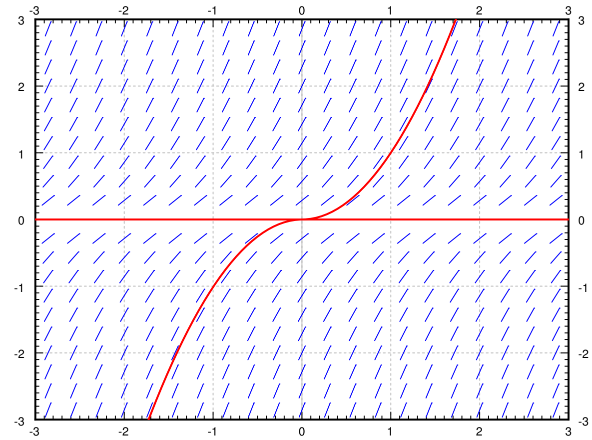


Figure 1.7: Slope field of $y' = 2\sqrt{|y|}$ with two solutions satisfying $y(0) = 0$.

Theorem 1.2.1 (Picard's theorem on existence and uniqueness). *If $f(x, y)$ is continuous (as a function of two variables) and $\frac{\partial f}{\partial y}$ exists and is continuous near some (x_0, y_0), then a solution to*

$$y' = f(x, y), \qquad y(x_0) = y_0,$$

exists (at least for x in some small interval) and is unique.

Note that the problems $y' = 1/x$, $y(0) = 0$ and $y' = 2\sqrt{|y|}$, $y(0) = 0$ do not satisfy the hypothesis of the theorem. Even if we can use the theorem, we ought to be careful about this existence business. It is quite possible that the solution only exists for a short while.

Example 1.2.3: For some constant A, solve:

$$y' = y^2, \qquad y(0) = A.$$

We know how to solve this equation. First assume that $A \neq 0$, so y is not equal to zero at least for some x near 0. So $x' = 1/y^2$, so $x = -1/y + C$, so $y = \frac{1}{C-x}$. If $y(0) = A$, then $C = 1/A$ so

$$y = \frac{1}{1/A - x}.$$

If $A = 0$, then $y = 0$ is a solution.

For example, when $A = 1$ the solution "blows up" at $x = 1$. Hence, the solution does not exist for all x even if the equation is nice everywhere. The equation $y' = y^2$ certainly looks nice.

For most of this course we will be interested in equations where existence and uniqueness holds, and in fact holds "globally" unlike for the equation $y' = y^2$.

1.2.3 Exercises

Exercise* 1.2.1*: *Sketch slope field for* $y' = e^{x-y}$. *How do the solutions behave as* x *grows? Can you guess a particular solution by looking at the slope field?*

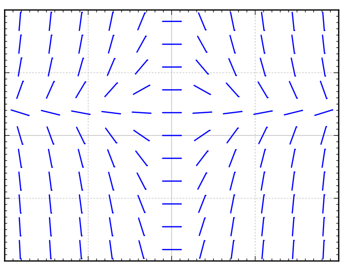
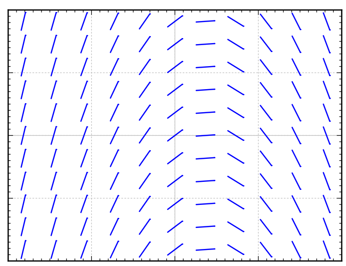
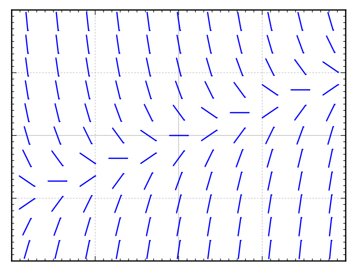
Exercise* 1.2.2*: *Sketch slope field for* $y' = x^2$.

Exercise* 1.2.3*: *Sketch slope field for* $y' = y^2$.

Exercise* 1.2.4*: *Is it possible to solve the equation* $y' = \frac{xy}{\cos x}$ *for* $y(0) = 1$? *Justify.*

Exercise* 1.2.5*: *Is it possible to solve the equation* $y' = y\sqrt{|x|}$ *for* $y(0) = 0$? *Is the solution unique? Justify.*

Exercise* 1.2.6*: *Match equations* $y' = 1 - x$, $y' = x - 2y$, $y' = x(1 - y)$ *to slope fields. Justify.*

a)  *b)*  *c)* 

Exercise* 1.2.7** (challenging):*** *Take* $y' = f(x, y)$, $y(0) = 0$, *where* $f(x, y) > 1$ *for all* x *and* y. *If the solution exists for all* x, *can you say what happens to* $y(x)$ *as* x *goes to positive infinity? Explain.*

Exercise* 1.2.8** (challenging):*** *Take* $(y - x)y' = 0$, $y(0) = 0$.

a) Find two distinct solutions.

b) Explain why this does not violate Picard's theorem.

Exercise* 1.2.9*: *Suppose* $y' = f(x, y)$. *What will the slope field look like, explain and sketch an example, if you know the following about* $f(x, y)$*:*

a) f *does not depend on* y.

b) f *does not depend on* x.

c) $f(t, t) = 0$ *for any number* t.

d) $f(x, 0) = 0$ *and* $f(x, 1) = 1$ *for all* x.

Exercise* 1.2.10*: *Find a solution to* $y' = |y|$, $y(0) = 0$. *Does Picard's theorem apply?*

Exercise* 1.2.11*: *Take an equation* $y' = (y - 2x)g(x, y) + 2$ *for some function* $g(x, y)$. *Can you solve the problem for the initial condition* $y(0) = 0$, *and if so what is the solution?*

Exercise* 1.2.12** (challenging):*** *Suppose $y' = f(x, y)$ is such that $f(x, 1) = 0$ for every x, f is continuous and $\frac{\partial f}{\partial y}$ exists and is continuous for every x and y.*

a) Guess a solution given the initial condition $y(0) = 1$.

b) Can graphs of two solutions of the equation for different initial conditions ever intersect?

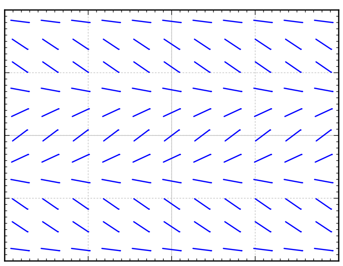
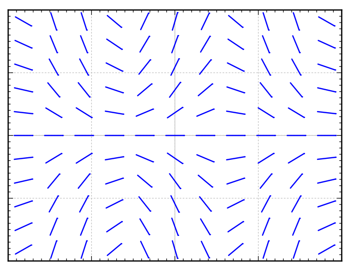
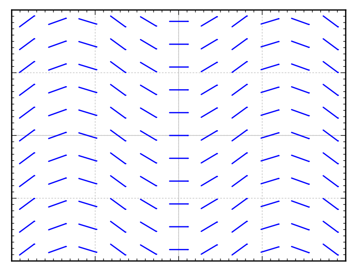
c) Given $y(0) = 0$, what can you say about the solution. In particular, can $y(x) > 1$ for any x? Can $y(x) = 1$ for any x? Why or why not?

Exercise* 1.2.101*: *Sketch the slope field of $y' = y^3$. Can you visually find the solution that satisfies $y(0) = 0$?*

Exercise* 1.2.102*: *Is it possible to solve $y' = xy$ for $y(0) = 0$? Is the solution unique?*

Exercise* 1.2.103*: *Is it possible to solve $y' = \frac{x}{x^2-1}$ for $y(1) = 0$?*

Exercise* 1.2.104*: *Match equations $y' = \sin x$, $y' = \cos y$, $y' = y\cos(x)$ to slope fields. Justify.*

a)  *b)*  *c)* 

Exercise* 1.2.105** (tricky):*** *Suppose*

$$f(y) = \begin{cases} 0 & \text{if } y > 0, \\ 1 & \text{if } y \leq 0. \end{cases}$$

Does $y' = f(y)$, $y(0) = 0$ have a continuously differentiable solution? Does Picard apply? Why, or why not?

Exercise* 1.2.106*: *Consider an equation of the form $y' = f(x)$ for some continuous function f, and an initial condition $y(x_0) = y_0$. Does a solution exist for all x? Why or why not?*

1.3 Separable equations

Note: 1 lecture, §1.4 in [EP], §2.2 in [BD]

When a differential equation is of the form $y' = f(x)$, we integrate: $y = \int f(x)\,dx + C$. Unfortunately, simply integrating no longer works for the general form of the equation $y' = f(x,y)$. Integrating both sides yields the rather unhelpful expression

$$y = \int f(x,y)\,dx + C.$$

Notice the dependence on y in the integral.

1.3.1 Separable equations

We say a differential equation is *separable* if we can write it as

$$y' = f(x)g(y),$$

for some functions $f(x)$ and $g(y)$. Let us write the equation in the Leibniz notation

$$\frac{dy}{dx} = f(x)g(y).$$

Then we rewrite the equation as

$$\frac{dy}{g(y)} = f(x)\,dx.$$

Both sides look like something we can integrate. We obtain

$$\int \frac{dy}{g(y)} = \int f(x)\,dx + C.$$

If we can find closed form expressions for these two integrals, we can, perhaps, solve for y.

Example 1.3.1: Take the equation

$$y' = xy.$$

Note that $y = 0$ is a solution. We will remember that fact and assume $y \neq 0$ from now on, so that we can divide by y. Write the equation as $\frac{dy}{dx} = xy$ or $\frac{dy}{y} = x\,dx$. Then

$$\int \frac{dy}{y} = \int x\,dx + C.$$

We compute the antiderivatives to get

$$\ln|y| = \frac{x^2}{2} + C,$$

or

$$|y| = e^{\frac{x^2}{2}+C} = e^{\frac{x^2}{2}} e^C = De^{\frac{x^2}{2}},$$

where $D > 0$ is some constant. Because $y = 0$ is also a solution and because of the absolute value, we can write:

$$y = De^{\frac{x^2}{2}},$$

for any number D (including zero or negative).

We check:

$$y' = Dxe^{\frac{x^2}{2}} = x\left(De^{\frac{x^2}{2}}\right) = xy.$$

Yay!

You may be worried that we integrated in two different variables. We seemingly did a different operation to each side. Perhaps we should be a little bit more careful and work through this method more rigorously. Consider

$$\frac{dy}{dx} = f(x)g(y).$$

We rewrite the equation as follows. Note that $y = y(x)$ is a function of x and so is $\frac{dy}{dx}$!

$$\frac{1}{g(y)}\frac{dy}{dx} = f(x).$$

We integrate both sides with respect to x:

$$\int \frac{1}{g(y)}\frac{dy}{dx}\,dx = \int f(x)\,dx + C.$$

We use the change of variables formula (substitution) on the left hand side:

$$\int \frac{1}{g(y)}\,dy = \int f(x)\,dx + C.$$

And we are done.

1.3.2 Implicit solutions

We sometimes get stuck even if we can do the integration. Consider the separable equation

$$y' = \frac{xy}{y^2+1}.$$

We separate variables,

$$\frac{y^2+1}{y}\,dy = \left(y + \frac{1}{y}\right)dy = x\,dx.$$

We integrate to get

$$\frac{y^2}{2} + \ln|y| = \frac{x^2}{2} + C,$$

or perhaps the less intimidating expression (where $D = 2C$)

$$y^2 + 2\ln|y| = x^2 + D.$$

It is not easy to find the solution explicitly—it is hard to solve for y. We, therefore, leave the solution in this form and call it an *implicit solution*. It is still easy to check that an implicit solution satisfies the differential equation. In this case, we differentiate with respect to x, and remember that y is a function of x, to get

$$y'\left(2y + \frac{2}{y}\right) = 2x.$$

Multiply both sides by y and divide by $2(y^2 + 1)$ and you will get exactly the differential equation. We leave this computation to the reader.

If you have an implicit solution, and you want to compute values for y, you might have to be tricky. You might get multiple solutions y for each x, so you have to pick one. Sometimes you can graph x as a function of y, and then turn your paper to see a graph. Sometimes you have to do more.

Computers are also good at some of these tricks. More advanced mathematical software usually has some way of plotting solutions to implicit equations. For example, for $D = 0$, if you plot all the points (x, y) that are solutions to $y^2 + 2\ln|y| = x^2$, you find the two curves in Figure 1.8 on the following page. This is not quite a graph of a function. For each x there are two choices of y. To find a function, you have to pick one of these two curves. You pick the one that satisfies your initial condition if you have one. For instance, the top curve satisfies the condition $y(1) = 1$. So for each D, we really got two solutions. As you can see, computing values from an implicit solution can be somewhat tricky. But sometimes, an implicit solution is the best we can do.

The equation above also has the solution $y = 0$. So the general solution is

$$y^2 + 2\ln|y| = x^2 + D, \qquad \text{and} \qquad y = 0.$$

Sometimes these extra solutions that came up due to division by zero such as $y = 0$ are called *singular solutions*.

1.3.3 Examples of separable equations

Example 1.3.2: Solve $x^2y' = 1 - x^2 + y^2 - x^2y^2$, $y(1) = 0$.

Factor the right-hand side

$$x^2y' = (1 - x^2)(1 + y^2).$$

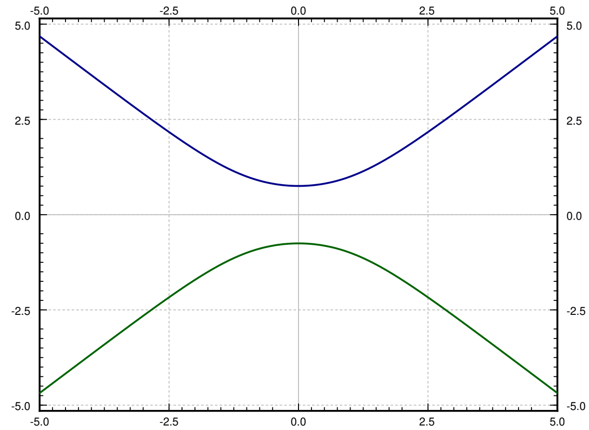


Figure 1.8: The implicit solution $y^2 + 2\ln|y| = x^2$ *to* $y' = \frac{xy}{y^2+1}$.

Separate variables, integrate, and solve for y:

$$\begin{aligned} \frac{y'}{1+y^2} &= \frac{1-x^2}{x^2}, \\ \frac{y'}{1+y^2} &= \frac{1}{x^2} - 1, \\ \arctan(y) &= \frac{-1}{x} - x + C, \\ y &= \tan\left(\frac{-1}{x} - x + C\right). \end{aligned}$$

Solve for the initial condition, $0 = \tan(-2 + C)$ to get $C = 2$ (or $C = 2 + \pi$, or $C = 2 + 2\pi$, etc.). The particular solution we seek is, therefore,

$$y = \tan\left(\frac{-1}{x} - x + 2\right).$$

Example 1.3.3: Bob made a cup of coffee, and Bob likes to drink coffee only once reaches 60 degrees Celsius and will not burn him. Initially at time $t = 0$ minutes, Bob measured the temperature and the coffee was 89 degrees Celsius. One minute later, Bob measured the coffee again and it had 85 degrees. The temperature of the room (the ambient temperature) is 22 degrees. When should Bob start drinking?

Let T be the temperature of the coffee in degrees Celsius, and let A be the ambient (room) temperature, also in degrees Celsius. Newton's law of cooling states that the rate at which the temperature of the coffee is changing is proportional to the difference between the ambient temperature and the temperature of the coffee. That is,

$$\frac{dT}{dt} = k(A - T),$$

for some positive constant k. For our setup $A = 22$, $T(0) = 89$, $T(1) = 85$. We separate variables and integrate (let C and D denote arbitrary constants):

$$\begin{aligned} \frac{1}{T-A}\,\frac{dT}{dt} &= -k, \\ \ln(T-A) &= -kt + C, \qquad \text{(note that } T - A > 0\text{)} \\ T - A &= D\,e^{-kt}, \\ T &= A + D\,e^{-kt}. \end{aligned}$$

That is, $T = 22 + D\,e^{-kt}$. We plug in the first condition: $89 = T(0) = 22 + D$, and hence $D = 67$. So $T = 22 + 67\,e^{-kt}$. The second condition says $85 = T(1) = 22 + 67\,e^{-k}$. Solving for k, we get $k = -\ln \frac{85-22}{67} \approx 0.0616$. Now we solve for the time t that gives us a temperature of 60 degrees. Namely, we solve

$$60 = 22 + 67e^{-0.0616t}$$

to get $t = -\frac{\ln \frac{60-22}{67}}{0.0616} \approx 9.21$ minutes. So Bob can begin to drink the coffee at just over 9 minutes from the time Bob made it. That is probably about the amount of time it took us to calculate how long it would take. See Figure 1.9.

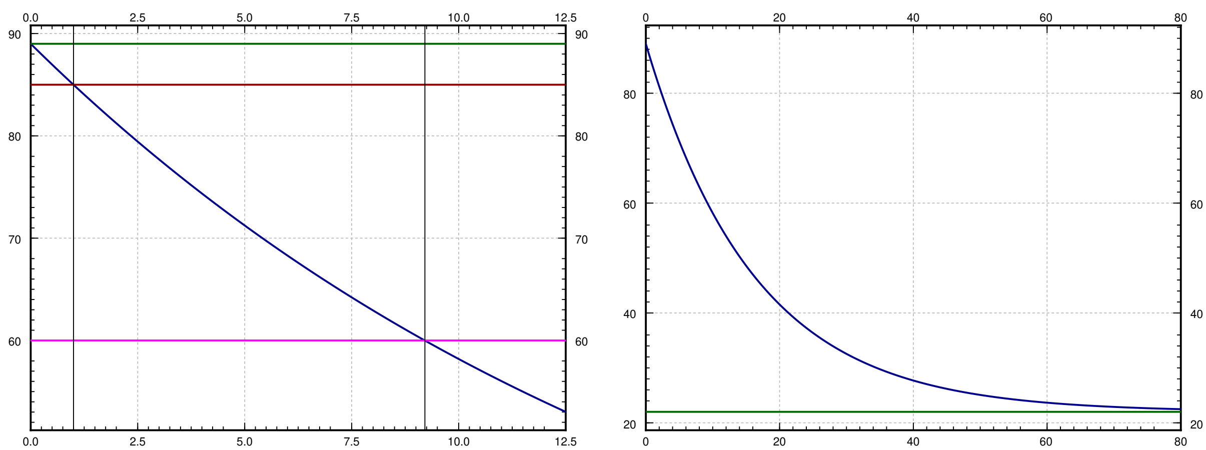


Figure 1.9: Graphs of the coffee temperature function $T(t)$. On the left, horizontal lines are drawn at temperatures 60, 85, and 89. Vertical lines are drawn at $t = 1$ and $t = 9.21$. Notice that the temperature of the coffee hits 85 at $t = 1$, and 60 at $t \approx 9.21$. On the right, the graph is over a longer period of time, with a horizontal line at the ambient temperature 22.

Example 1.3.4: Find the general solution to $y' = \frac{-xy^2}{3}$ (including any singular solutions).

First note that $y = 0$ is a solution (a singular solution). Now assume that $y \neq 0$.

$$\frac{-3}{y^2}y' = x,$$

$$\frac{3}{y} = \frac{x^2}{2} + C,$$
$$y = \frac{3}{x^2/2 + C} = \frac{6}{x^2 + 2C}.$$

So the general solution is

$$y = \frac{6}{x^2 + 2C} \qquad \text{and} \qquad y = 0.$$

1.3.4 Exercises

Exercise* 1.3.1*: *Solve* $y' = x/y$.

Exercise* 1.3.2*: *Solve* $y' = x^2 y$.

Exercise* 1.3.3*: *Solve* $\frac{dx}{dt} = (x^2 - 1)\, t$, *for* $x(0) = 0$.

Exercise* 1.3.4*: *Solve* $\frac{dx}{dt} = x \sin(t)$, *for* $x(0) = 1$.

Exercise* 1.3.5*: *Solve* $\frac{dy}{dx} = xy + x + y + 1$. *Hint: Factor the right-hand side.*

Exercise* 1.3.6*: *Solve* $xy' = y + 2x^2 y$, *where* $y(1) = 1$.

Exercise* 1.3.7*: *Solve* $\frac{dy}{dx} = \frac{y^2+1}{x^2+1}$, *for* $y(0) = 1$.

Exercise* 1.3.8*: *Find an implicit solution for* $\frac{dy}{dx} = \frac{x^2+1}{y^2+1}$, *for* $y(0) = 1$.

Exercise* 1.3.9*: *Find an explicit solution for* $y' = xe^{-y}$, $y(0) = 1$.

Exercise* 1.3.10*: *Find an explicit solution for* $xy' = e^{-y}$, *for* $y(1) = 1$.

Exercise* 1.3.11*: *Find an explicit solution for* $y' = ye^{-x^2}$, $y(0) = 1$. *It is alright to leave a definite integral in your answer.*

Exercise* 1.3.12*: *Suppose a cup of coffee is at 100 degrees Celsius at time* $t = 0$, *it is at 70 degrees at* $t = 10$ *minutes, and it is at 50 degrees at* $t = 20$ *minutes. Compute the ambient temperature.*

Exercise* 1.3.101*: *Solve* $y' = 2xy$.

Exercise* 1.3.102*: *Solve* $x' = 3xt^2 - 3t^2$, $x(0) = 2$.

Exercise* 1.3.103*: *Find an implicit solution for* $x' = \frac{1}{3x^2+1}$, $x(0) = 1$.

Exercise* 1.3.104*: *Find an explicit solution to* $xy' = y^2$, $y(1) = 1$.

Exercise* 1.3.105*: *Find an implicit solution to* $y' = \frac{\sin(x)}{\cos(y)}$.

Exercise* 1.3.106*: *Take Example 1.3.3 with the same numbers: 89 degrees at* $t = 0$, *85 degrees at* $t = 1$, *and ambient temperature of 22 degrees. Suppose these temperatures were measured with precision of* ± 0.5 *degrees. Given this imprecision, the time it takes the coffee to cool to (exactly) 60 degrees is also only known in a certain range. Find this range. Hint: Think about what kind of error makes the cooling time longer and what shorter.*

Exercise* 1.3.107*: *A population* x *of rabbits on an island is modeled by* $x' = x - (1/1000)x^2$, *where the independent variable is time in months. At time* $t = 0$, *there are 40 rabbits on the island.*

a) *Find the solution to the equation with the initial condition.*
b) *How many rabbits are on the island in 1 month, 5 months, 10 months, 15 months (round to the nearest integer).*

1.4 Linear equations and the integrating factor

Note: 1 lecture, §1.5 in [EP], §2.1 in [BD]

One of the most important types of equations we will learn to solve are the so-called *linear equations*. In fact, the majority of the course is about linear equations. In this section we focus on the *first order linear equation*. A first order equation is linear if we can put it into the form:

$$y' + p(x)y = f(x). \tag{1.3}$$

The word "linear" means linear in y and y'; no higher powers nor functions of y or y' appear. The dependence on x can be more complicated.

Solutions of linear equations have nice properties. For example, the solution exists wherever $p(x)$ and $f(x)$ are defined, and has essentially the same regularity (read: it is just as nice). But most importantly for us right now, there is a method for solving linear first order equations.

The trick is to rewrite the left-hand side of (1.3) as a derivative of a product of y with another function. To this end, we find a function $r(x)$ such that

$$r(x)y' + r(x)p(x)y = \frac{d}{dx}\Big[r(x)y\Big].$$

This is the left-hand side of (1.3) multiplied by $r(x)$. If we multiply (1.3) by $r(x)$, we obtain

$$\frac{d}{dx}\Big[r(x)y\Big] = r(x)f(x).$$

We can now integrate both sides, which we can do as the right-hand side does not depend on y and the left-hand side is written as a derivative of a function. After the integration, we solve for y by dividing by $r(x)$. The function $r(x)$ is called the *integrating factor* and the method is called the *integrating factor method*.

We are looking for a function $r(x)$, such that if we differentiate it, we get the same function back multiplied by $p(x)$. That seems like a job for the exponential function! Let

$$r(x) = e^{\int p(x)\,dx}.$$

We compute:

$$\begin{aligned}
y' + p(x)y &= f(x), \\
e^{\int p(x)\,dx}y' + e^{\int p(x)\,dx}p(x)y &= e^{\int p(x)\,dx}f(x), \\
\frac{d}{dx}\left[e^{\int p(x)\,dx}y\right] &= e^{\int p(x)\,dx}f(x), \\
e^{\int p(x)\,dx}y &= \int e^{\int p(x)\,dx}f(x)\,dx + C, \\
y &= e^{-\int p(x)\,dx}\left(\int e^{\int p(x)\,dx}f(x)\,dx + C\right).
\end{aligned}$$

Of course, to get a closed form formula for y, we need to be able to find a closed form formula for the integrals appearing above.

Example 1.4.1: Solve

$$y' + 2xy = e^{x-x^2}, \qquad y(0) = -1.$$

First note that $p(x) = 2x$ and $f(x) = e^{x-x^2}$. The integrating factor is $r(x) = e^{\int p(x)\,dx} = e^{x^2}$. We multiply both sides of the equation by $r(x)$ to get

$$\begin{aligned} e^{x^2}y' + 2xe^{x^2}y &= e^{x-x^2}e^{x^2}, \\ \frac{d}{dx}\left[e^{x^2}y\right] &= e^x. \end{aligned}$$

We integrate

$$\begin{aligned} e^{x^2}y &= e^x + C, \\ y &= e^{x-x^2} + Ce^{-x^2}. \end{aligned}$$

Next, we solve for the initial condition $-1 = y(0) = 1 + C$, so $C = -2$. The solution is

$$y = e^{x-x^2} - 2e^{-x^2}.$$

Note that we do not care which antiderivative we take when computing $e^{\int p(x)dx}$. You can always add a constant of integration, but those constants will not matter in the end.

***Exercise* 1.4.1:** *Try it! Add a constant of integration to the integral in the integrating factor and show that the solution you get in the end is the same as what we got above.*

Advice: Do not try to remember the formula for y itself, that is way too hard. It is easier to remember the process and repeat it.

Since we cannot always evaluate the integrals in closed form, it is useful to know how to write the solution in definite integral form. A definite integral is something that you can plug into a computer or a calculator. Suppose we are given

$$y' + p(x)y = f(x), \qquad y(x_0) = y_0.$$

Look at the solution and write the integrals as definite integrals.

$$\boxed{y(x) = e^{-\int_{x_0}^{x} p(s)\,ds}\left(\int_{x_0}^{x} e^{\int_{x_0}^{t} p(s)\,ds} f(t)\,dt + y_0\right).} \tag{1.4}$$

You should be careful to properly use dummy variables here. If you now plug such a formula into a computer or a calculator, it will be happy to give you numerical answers.

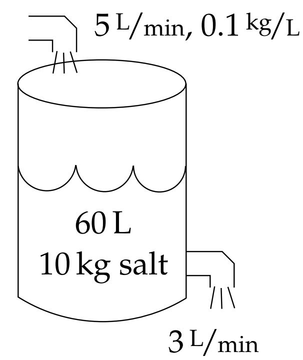
***Exercise* 1.4.2:** *Check that $y(x_0) = y_0$ in formula (1.4).*

***Exercise* 1.4.3:** *Write the solution of the following problem as a definite integral, but try to simplify as far as you can. You will not be able to find the solution in closed form.*

$$y' + y = e^{x^2-x}, \qquad y(0) = 10.$$

Remark 1.4.1: Before we move on, we should note some interesting properties of linear equations. First, for the linear initial value problem $y' + p(x)y = f(x)$, $y(x_0) = y_0$, there is an explicit formula (1.4) for the solution. Second, it follows from the formula (1.4) that if $p(x)$ and $f(x)$ are continuous on some interval (a, b), then the solution $y(x)$ exists and is differentiable on (a, b). Compare with the simple nonlinear example we have seen previously, $y' = y^2$, and compare to Theorem 1.2.1.

Example 1.4.2: Let us discuss a common simple application of linear equations. Real life applications of this type of problem include figuring out the concentration of chemicals in bodies of water (rivers and lakes).



A 100 liter tank contains 10 kilograms of salt dissolved in 60 liters of water. Solution of water and salt (brine) with concentration of 0.1 kilograms per liter is flowing in at the rate of 5 liters a minute. The solution in the tank is well stirred and flows out at a rate of 3 liters a minute. How much salt is in the tank when the tank is full?

Let us come up with the equation. Let x denote the kilograms of salt in the tank, let t denote the time in minutes. For a small change Δt in time, the change in x (denoted Δx) is approximately

$$\Delta x \approx (\text{rate in} \times \text{concentration in})\Delta t - (\text{rate out} \times \text{concentration out})\Delta t.$$

Dividing through by Δt and taking the limit $\Delta t \to 0$, we see that

$$\frac{dx}{dt} = (\text{rate in} \times \text{concentration in}) - (\text{rate out} \times \text{concentration out}).$$

In our example,

$$\begin{aligned} \text{rate in} &= 5, \\ \text{concentration in} &= 0.1, \\ \text{rate out} &= 3, \\ \text{concentration out} &= \frac{x}{\text{volume}} = \frac{x}{60 + (5-3)t}. \end{aligned}$$

Our equation is, therefore,

$$\frac{dx}{dt} = (5 \times 0.1) - \left(3 \frac{x}{60 + 2t}\right).$$

Or in the form (1.3),

$$\frac{dx}{dt} + \frac{3}{60 + 2t} x = 0.5.$$

Let us solve. The integrating factor is

$$r(t) = \exp\left(\int \frac{3}{60+2t}dt\right) = \exp\left(\frac{3}{2}\ln(60+2t)\right) = (60+2t)^{3/2}.$$

We multiply both sides of the equation to get

$$\begin{aligned}
(60+2t)^{3/2}\frac{dx}{dt} + (60+2t)^{3/2}\frac{3}{60+2t}x &= 0.5(60+2t)^{3/2},\\
\frac{d}{dt}\left[(60+2t)^{3/2}x\right] &= 0.5(60+2t)^{3/2},\\
(60+2t)^{3/2}x &= \int 0.5(60+2t)^{3/2}dt + C,\\
x &= (60+2t)^{-3/2}\int \frac{(60+2t)^{3/2}}{2}dt + C(60+2t)^{-3/2},\\
x &= (60+2t)^{-3/2}\frac{1}{10}(60+2t)^{5/2} + C(60+2t)^{-3/2},\\
x &= \frac{60+2t}{10} + C(60+2t)^{-3/2}.
\end{aligned}$$

To find C, note that at $t = 0$, we have $x = 10$. That is,

$$10 = x(0) = \frac{60}{10} + C(60)^{-3/2} = 6 + C(60)^{-3/2},$$

or

$$C = 4(60^{3/2}) \approx 1859.03.$$

We know 5 liters per minute are flowing in and 3 liters per minute are flowing out, so the volume is increasing by 2 liters a minute. So the tank is full when $60 + 2t = 100$, or when $t = 20$. We are interested in the value of x when the tank is full, that is we want to compute $x(20)$:

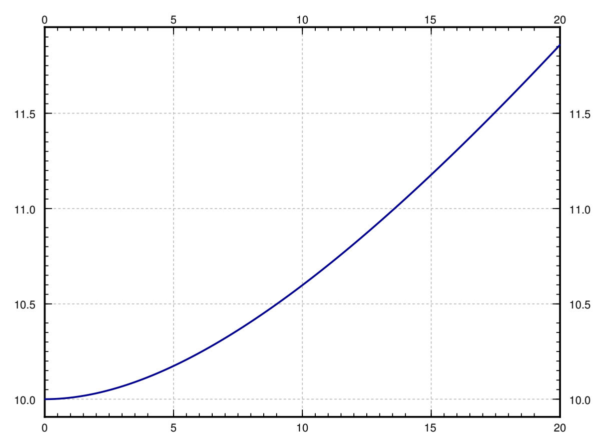


Figure 1.10: Graph of the solution x kilograms of salt in the tank at time t.

$$\begin{aligned}
x(20) &= \frac{60+40}{10} + C(60+40)^{-3/2}\\
&\approx 10 + 1859.03(100)^{-3/2} \approx 11.86.
\end{aligned}$$

There are 11.86 kg of salt in the tank when it is full. See Figure 1.10 for the graph of x over t.

The concentration when the tank is full is approximately $11.86/100 = 0.1186$ kg/liter, and we started with $1/6$ or approximately 0.1667 kg/liter.

1.4.1 Exercises

In the exercises, feel free to leave answer as a definite integral if a closed form solution cannot be found. If you can find a closed form solution, you should give that.

***Exercise* 1.4.4:** *Solve* $y' + xy = x$.

***Exercise* 1.4.5:** *Solve* $y' + 6y = e^x$.

***Exercise* 1.4.6:** *Solve* $y' + 3x^2y = \sin(x)\, e^{-x^3}$, *with* $y(0) = 1$.

***Exercise* 1.4.7:** *Solve* $y' + \cos(x)y = \cos(x)$.

***Exercise* 1.4.8:** *Solve* $\frac{1}{x^2+1}\, y' + xy = 3$, *with* $y(0) = 0$.

***Exercise* 1.4.9:** *Suppose there are two lakes located on a stream. Clean water flows into the first lake, then the water from the first lake flows into the second lake, and then water from the second lake flows further downstream. The in and out flow from each lake is 500 liters per hour. The first lake contains 100 thousand liters of water and the second lake contains 200 thousand liters of water. A truck with 500 kg of toxic substance crashes into the first lake. Assume that the water is being continually mixed perfectly by the stream.*

a) Find the concentration of toxic substance as a function of time in both lakes.

b) When will the concentration in the first lake be below 0.001 kg per liter?

c) When will the concentration in the second lake be maximal?

***Exercise* 1.4.10:** *Newton's law of cooling states that* $\frac{dx}{dt} = -k(x - A)$ *where* x *is the temperature,* t *is time,* A *is the ambient temperature, and* $k > 0$ *is a constant. Suppose that* $A = A_0 \cos(\omega t)$ *for some constants* A_0 *and* ω. *That is, the ambient temperature oscillates (for example night and day temperatures).*

a) Find the general solution.

b) In the long term, will the initial conditions make much of a difference? Why or why not?

***Exercise* 1.4.11:** *Initially 5 grams of salt are dissolved in 20 liters of water. Brine with concentration of salt 2 grams of salt per liter is added at a rate of 3 liters a minute. The tank is mixed well and is drained at 3 liters a minute. How long does the process have to continue until there are 20 grams of salt in the tank?*

***Exercise* 1.4.12:** *Initially a tank contains 10 liters of pure water. Brine of unknown (but constant) concentration of salt is flowing in at 1 liter per minute. The water is mixed well and drained at 1 liter per minute. In 20 minutes there are 15 grams of salt in the tank. What is the concentration of salt in the incoming brine?*

***Exercise* 1.4.101:** *Solve* $y' + 3x^2y = x^2$.

***Exercise* 1.4.102:** *Solve* $y' + 2\sin(2x)y = 2\sin(2x)$, $y(\pi/2) = 3$.

Exercise* 1.4.103*: *Suppose a water tank is being pumped out at* $3\,{}^{\mathrm{L}}\!/\!_{\mathrm{min}}$*. The water tank starts at 10 L of clean water. Water with toxic substance is flowing into the tank at* $2\,{}^{\mathrm{L}}\!/\!_{\mathrm{min}}$*, with concentration* $20t\,{}^{\mathrm{g}}\!/\!_{\mathrm{L}}$ *at time* t*. When the tank is half empty, how many grams of toxic substance are in the tank (assuming perfect mixing)?*

Exercise* 1.4.104*: *There is bacteria on a plate and a toxic substance is being added that slows down the rate of growth of the bacteria. That is, suppose that* $\frac{dP}{dt} = (2 - 0.1\,t)P$*. If* $P(0) = 1000$*, find the population at* $t = 5$*.*

Exercise* 1.4.105*: *A cylindrical water tank has water flowing in at* I *cubic meters per second. Let* A *be the area of the cross section of the tank in square meters. Suppose water is flowing out from the bottom of the tank at a rate proportional to the height of the water level. Set up the differential equation for* h*, the height of the water, introducing and naming constants that you need. You should also give the units for your constants.*

1.5 Substitution

Note: 1 lecture, can safely be skipped, §1.6 in [EP], not in [BD]

Just as when solving integrals, one method to try is to change variables to end up with a simpler equation to solve.

1.5.1 Substitution

The equation
$$y' = (x - y + 1)^2$$
is neither separable nor linear. What can we do? How about trying to change variables, so that in the new variables the equation is simpler. We use another variable v, which we treat as a function of x. We try
$$v = x - y + 1.$$
We need to figure out y' in terms of v', v and x. We differentiate (in x) to obtain $v' = 1 - y'$. So $y' = 1 - v'$. We plug this into the equation to get
$$1 - v' = v^2.$$
In other words, $v' = 1 - v^2$. Such an equation we know how to solve by separating variables:
$$\frac{1}{1 - v^2}\, dv = dx.$$
So
$$\frac{1}{2} \ln \left| \frac{v+1}{v-1} \right| = x + C, \qquad \text{or} \qquad \left| \frac{v+1}{v-1} \right| = e^{2x+2C}, \qquad \text{or} \qquad \frac{v+1}{v-1} = De^{2x},$$
for some constant D. Note that $v = 1$ and $v = -1$ are also solutions.

Now we need to "unsubstitute" to obtain
$$\frac{x - y + 2}{x - y} = De^{2x},$$
and also the two solutions $x - y + 1 = 1$ or $y = x$, and $x - y + 1 = -1$ or $y = x + 2$. We solve the first equation for y:
$$\begin{aligned}
x - y + 2 &= (x - y)De^{2x}, \\
x - y + 2 &= Dxe^{2x} - yDe^{2x}, \\
-y + yDe^{2x} &= Dxe^{2x} - x - 2, \\
y\left(-1 + De^{2x}\right) &= Dxe^{2x} - x - 2, \\
y &= \frac{Dxe^{2x} - x - 2}{De^{2x} - 1}.
\end{aligned}$$
Note that $D = 0$ gives $y = x + 2$, but no value of D gives the solution $y = x$.

Substitution in differential equations is applied in much the same way that it is applied in calculus. You guess. Several different substitutions might work. There are some general patterns to look for. We summarize a few of these in a table.

When you see	Try substituting
yy'	$v = y^2$
y^2y'	$v = y^3$
$(\cos y)y'$	$v = \sin y$
$(\sin y)y'$	$v = \cos y$
$e^y y'$	$v = e^y$

Usually you try to substitute in the "most complicated" part of the equation with the hopes of simplifying it. The table above is just a rule of thumb. You might have to modify your guesses. If a substitution does not work (it does not make the equation any simpler), try a different one.

1.5.2 Bernoulli equations

There are some forms of equations where there is a general rule for substitution that always works. One such example is the so-called *Bernoulli equation**:

$$y' + p(x)y = q(x)y^n.$$

This equation looks a lot like a linear equation except for the y^n. If $n = 0$ or $n = 1$, then the equation is linear and we can solve it. Otherwise, the substitution $v = y^{1-n}$ transforms the Bernoulli equation into a linear equation. Note that n need not be an integer.

Example 1.5.1: Solve

$$xy' + y(x+1) + xy^5 = 0, \qquad y(1) = 1.$$

The equation is a Bernoulli equation, $p(x) = (x+1)/x$ and $q(x) = -1$. We substitute

$$v = y^{1-5} = y^{-4}, \qquad v' = -4y^{-5}y'.$$

In other words, $(-1/4)\, y^5 v' = y'$. So

$$\begin{aligned} xy' + y(x+1) + xy^5 &= 0, \\ \frac{-xy^5}{4}v' + y(x+1) + xy^5 &= 0, \\ \frac{-x}{4}v' + y^{-4}(x+1) + x &= 0, \end{aligned}$$

*There are several things called Bernoulli equations, this is just one of them. The Bernoullis were a prominent Swiss family of mathematicians. These particular equations are named for Jacob Bernoulli (1654–1705).

$$\frac{-x}{4}v' + v(x+1) + x = 0,$$

and finally

$$v' - \frac{4(x+1)}{x}v = 4.$$

The equation is now linear. We can use the integrating factor method. In particular, we use formula (1.4). We assume that $x > 0$ so $|x| = x$. This assumption is OK, as our initial condition is at $x = 1 > 0$. Let us compute the integrating factor. Here $p(s)$ from formula (1.4) is $\frac{-4(s+1)}{s}$.

$$\begin{aligned} e^{\int_1^x p(s)\,ds} &= \exp\left(\int_1^x \frac{-4(s+1)}{s}\,ds\right) = e^{-4x-4\ln(x)+4} = e^{-4x+4}x^{-4} = \frac{e^{-4x+4}}{x^4}, \\ e^{-\int_1^x p(s)\,ds} &= e^{4x+4\ln(x)-4} = e^{4x-4}x^4. \end{aligned}$$

We now plug in to (1.4)

$$\begin{aligned} v(x) &= e^{-\int_1^x p(s)\,ds}\left(\int_1^x e^{\int_1^t p(s)\,ds}4\,dt + 1\right) \\ &= e^{4x-4}x^4\left(\int_1^x 4\frac{e^{-4t+4}}{t^4}\,dt + 1\right). \end{aligned}$$

The integral in this expression is not possible to find in closed form. As we said before, it is perfectly fine to have a definite integral in our solution. Now "unsubstitute"

$$\begin{aligned} y^{-4} &= e^{4x-4}x^4\left(4\int_1^x \frac{e^{-4t+4}}{t^4}\,dt + 1\right), \\ y &= \frac{e^{-x+1}}{x\left(4\int_1^x \frac{e^{-4t+4}}{t^4}\,dt + 1\right)^{1/4}}. \end{aligned}$$

1.5.3 Homogeneous equations

Another type of equations we can solve by substitution are the so-called *homogeneous equations*. Suppose that we can write the differential equation as

$$y' = F\left(\frac{y}{x}\right).$$

Here we try the substitutions

$$v = \frac{y}{x} \qquad \text{and therefore} \qquad y' = v + xv'.$$

We note that the equation is transformed into

$$v + xv' = F(v) \qquad \text{or} \qquad xv' = F(v) - v \qquad \text{or} \qquad \frac{v'}{F(v)-v} = \frac{1}{x}.$$

Hence an implicit solution is

$$\int \frac{1}{F(v) - v} \, dv = \ln |x| + C.$$

Clearly this solution does not work when $x = 0$ (we would, afterall, divide by zero in ${}^{y}/_{x}$). So we will either assume $x > 0$ or $x < 0$ depending on the initial condition.

Example 1.5.2: Solve

$$x^2 y' = y^2 + xy, \qquad y(1) = 1.$$

We put the equation into the form $y' = ({}^{y}/_{x})^2 + {}^{y}/_{x}$, that is, $F(v) = v^2 + v$. As the initial condition is for a positive x value, we will assume $x > 0$. We substitute $v = {}^{y}/_{x}$ to get the separable equation

$$xv' = v^2 + v - v = v^2,$$

which has a solution

$$\begin{aligned} \int \frac{1}{v^2} \, dv &= \ln |x| + C, \\ \frac{-1}{v} &= \ln x + C, \\ v &= \frac{-1}{\ln x + C}. \end{aligned}$$

We unsubstitute

$$\frac{y}{x} = \frac{-1}{\ln x + C}, \qquad \text{or} \qquad y = \frac{-x}{\ln x + C}.$$

We want $y(1) = 1$, so

$$1 = y(1) = \frac{-1}{\ln 1 + C} = \frac{-1}{C}.$$

Thus $C = -1$ and the solution we are looking for is

$$y = \frac{-x}{\ln x - 1}.$$

1.5.4 Exercises

Hint: Answers need not always be in closed form.

***Exercise* 1.5.1:** *Solve $y' + y(x^2 - 1) + xy^6 = 0$, with $y(1) = 1$.*

***Exercise* 1.5.2:** *Solve $2yy' + 1 = y^2 + x$, with $y(0) = 1$.*

***Exercise* 1.5.3:** *Solve $y' + xy = y^4$, with $y(0) = 1$.*

***Exercise* 1.5.4:** *Solve $yy' + x = \sqrt{x^2 + y^2}$.*

***Exercise* 1.5.5:** *Solve $y' = (x + y - 1)^2$.*

Exercise* 1.5.6*: *Solve* $y' = \frac{x^2-y^2}{xy}$*, with* $y(1) = 2$.

Exercise* 1.5.101*: *Solve* $xy' + y + y^2 = 0$, $y(1) = 2$.

Exercise* 1.5.102*: *Solve* $xy' + y + x = 0$, $y(1) = 1$.

Exercise* 1.5.103*: *Solve* $y^2y' = y^3 - 3x$, $y(0) = 2$.

Exercise* 1.5.104*: *Solve* $2yy' = e^{y^2-x^2} + 2x$.

1.6 Autonomous equations

Note: 1 lecture, §2.2 in [EP], §2.5 in [BD]

Consider problems of the form

$$\frac{dx}{dt} = f(x),$$

where the derivative of solutions depends only on x (the dependent variable). Such equations are called *autonomous equations*. If we think of t as time, the naming comes from the fact that the equation is independent of time.

We return to the cooling coffee problem (Example 1.3.3). Newton's law of cooling says

$$\frac{dx}{dt} = k(A - x),$$

where x is the temperature, t is time, k is some positive constant, and A is the ambient temperature. See Figure 1.11 for an example with $k = 0.3$ and $A = 5$.

Note the solution $x = A$ (in the figure $x = 5$). We call these constant solutions the *equilibrium solutions*. The points on the x-axis where $f(x) = 0$ are called *critical points*. The point $x = A$ is a critical point. In fact, each critical point corresponds to an equilibrium solution. Note also, by looking at the graph, that the solution $x = A$ is "stable" in that small perturbations in x do not lead to substantially different solutions as t grows. If we change the initial condition a little bit, then as $t \to \infty$ we get $x(t) \to A$. We call such a critical point *stable*. In this simple example it turns out that all solutions in fact go to A as $t \to \infty$. If a critical point is not stable, we say it is *unstable*.

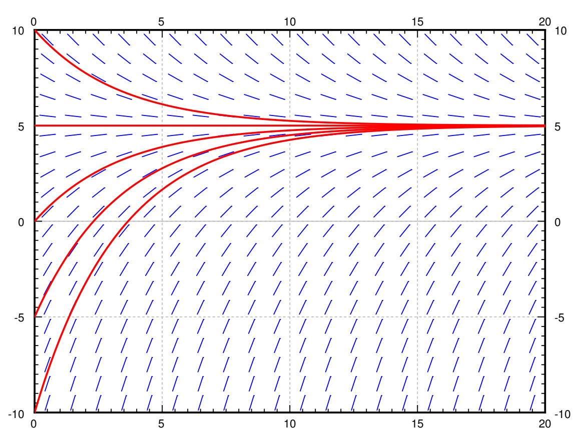


Figure 1.11: The slope field and some solutions of $x' = 0.3\,(5 - x)$.

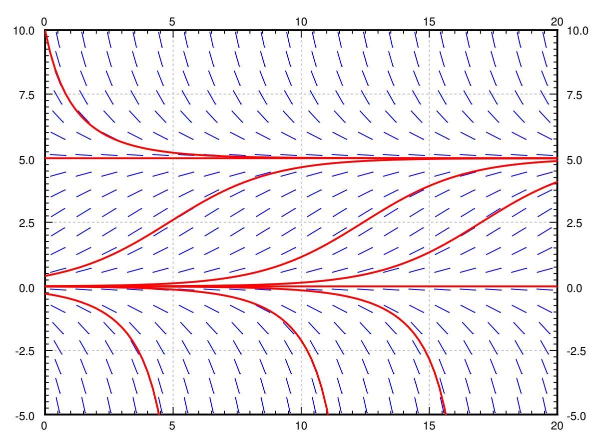


Figure 1.12: The slope field and some solutions of $x' = 0.1\,x\,(5 - x)$.

Consider now the *logistic equation*

$$\frac{dx}{dt} = kx(M - x),$$

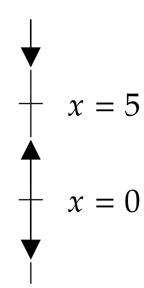
for some positive k and M. This equation is commonly used to model population if we know the limiting population M, that is the maximum sustainable population. The logistic equation leads to less catastrophic predictions on world population than $x' = kx$. In the real world there is no such thing as negative population, but we will still consider negative x for the purposes of the math.

See Figure 1.12 on the preceding page for an example, $x' = 0.1x(5 - x)$. There are two critical points, $x = 0$ and $x = 5$. The critical point at $x = 5$ is stable, while the critical point at $x = 0$ is unstable. It is not necessary to find the exact solutions to understand their long term behavior, that is, behavior as time goes to infinity. From the slope field above of $x' = 0.1x(5 - x)$, we see that

$$\lim_{t\to\infty} x(t) = \begin{cases} 5 & \text{if } \; x(0) > 0, \\ 0 & \text{if } \; x(0) = 0, \\ \text{DNE or } -\infty & \text{if } \; x(0) < 0. \end{cases}$$

Here DNE means "does not exist." From just looking at the slope field we cannot quite decide what happens if $x(0) < 0$. It could be that the solution does not exist for t all the way to ∞. Think of the equation $x' = x^2$; we have seen that solutions only exist for some finite period of time. Same can happen here. In our example equation above it turns out that the solution does not exist for all time, but to see that we would have to solve the equation. In any case, the solution does go to $-\infty$, but it may get there rather quickly.

If we are interested only in the long term behavior of the solution, we would be doing unnecessary work if we solved the equation exactly. We could draw the slope field, but it is easier to just look at the *phase diagram* or *phase portrait*, which is a simple way to visualize the behavior of autonomous equations. In this case there is one dependent variable x. We draw the x-axis, we mark all the critical points, and then we draw arrows in between. Since x is the dependent variable we draw the axis vertically, as it appears in the slope field diagrams above. If $f(x) > 0$, we draw an up arrow. If $f(x) < 0$, we draw a down arrow. To figure this out, we could just plug in some x between the critical points, $f(x)$ will have the same sign at all x between two critical points as long $f(x)$ is continuous. For example, $f(6) = -0.6 < 0$, so $f(x) < 0$ for $x > 5$, and the arrow above $x = 5$ is a down arrow. Next, $f(1) = 0.4 > 0$, so $f(x) > 0$ whenever $0 < x < 5$, and the arrow points up. Finally, $f(-1) = -0.6 < 0$ so $f(x) < 0$ when $x < 0$, and the arrow points down.



Armed with the phase diagram, it is easy to sketch the solutions approximately: As time t moves from left to right, the graph of a solution goes up if the arrow is up, and it goes down if the arrow is down.

***Exercise* 1.6.1:** *Try sketching a few solutions simply from looking at the phase diagram. Check with the preceding graphs if you are getting the same type of curves.*

Once we draw the phase diagram, we classify critical points as stable or unstable*. Since any mathematical model we cook up will only be an approximation to the real world, unstable points are generally bad news.

We remark that you can figure out the arrows by plotting the graph $y = f(x)$. However, in that case note that x is then the dependent variable and will be on the horizontal axis.

Let us think about the logistic equation with harvesting. Suppose an alien race really likes to eat humans. They keep a planet with humans and harvest the humans at a rate of h million humans per year. Suppose x is the number of humans in millions on the planet and t is time in years. Let M be the limiting population when no harvesting is done. The number $k > 0$ is a constant depending on how fast humans multiply. Our equation becomes

$$\frac{dx}{dt} = kx(M - x) - h.$$

We expand the right-hand side and set it to zero.

$$kx(M - x) - h = -kx^2 + kMx - h = 0.$$

Solving for the critical points, let us call them A and B, we get

$$A = \frac{kM + \sqrt{(kM)^2 - 4hk}}{2k}, \qquad B = \frac{kM - \sqrt{(kM)^2 - 4hk}}{2k}.$$

***Exercise* 1.6.2:** *Sketch a phase diagram for different possibilities. Note that these possibilities are $A > B$, or $A = B$, or A and B both complex (i.e. no real solutions). Hint: Fix some simple k and M and then vary h.*

For example, let $M = 8$ and $k = 0.1$. When $h = 1$, then A and B are distinct and positive. See Figure 1.13 on the next page for the slope field. As long as the population starts above B, which is approximately 1.55 million, then the population will not die out, it will tend towards $A \approx 6.45$ million. If ever a catastrophe happens and the population drops below B, humans will die out, and the fast food restaurant serving them will go out of business.

*Unstable points with one of the arrows pointing towards the critical point are sometimes called *semistable*.

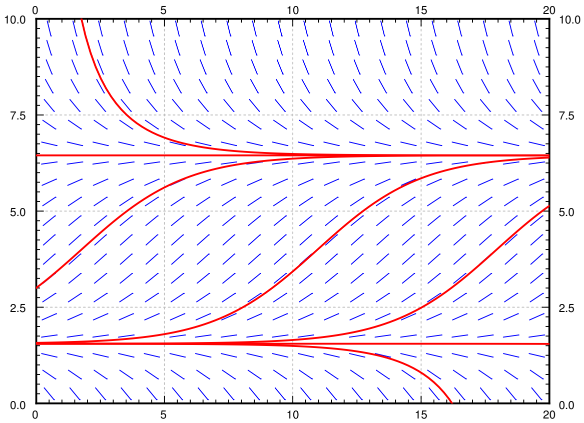


Figure 1.13: The slope field and some solutions of $x' = 0.1\,x\,(8 - x) - 1$.

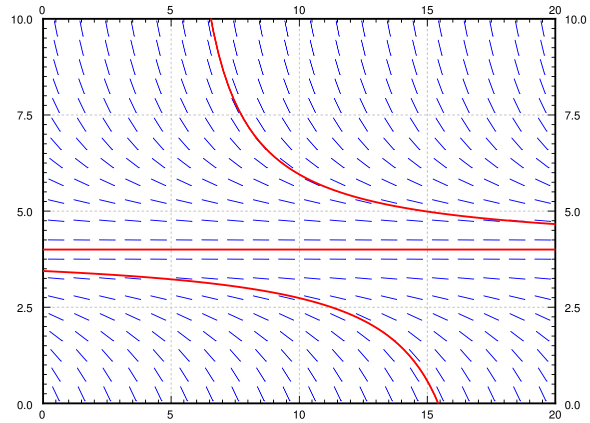


Figure 1.14: The slope field and some solutions of $x' = 0.1\,x\,(8 - x) - 1.6$.

When $h = 1.6$, then $A = B = 4$. There is only one critical point and it is unstable. When the population starts above 4 million, it will tend towards 4 million. However, if it ever drops below 4 million, perhaps a worse than normal hurricane season one year, then humans will die out on the planet. This scenario is not one that we (as the human fast food proprietor) want to be in. A small perturbation of the equilibrium state and we are out of business. There is no room for error. See Figure 1.14.

Finally, if we are harvesting at 2 million humans per year, there are no critical points. The population will always plummet towards zero, no matter how well stocked the planet starts. See Figure 1.15.

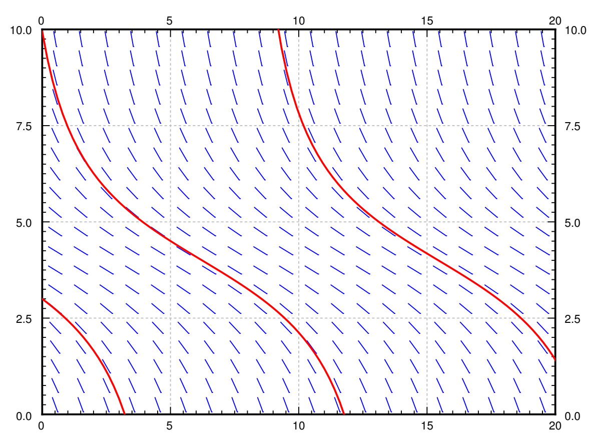


Figure 1.15: The slope field and some solutions of $x' = 0.1\,x\,(8 - x) - 2$.

1.6.1 Exercises

***Exercise* 1.6.3:** *Consider* $x' = x^2$.

a) *Draw the phase diagram, find the critical points, and mark them stable or unstable.*

b) *Sketch typical solutions of the equation.*

c) *Find* $\lim_{t\to\infty} x(t)$ *for the solution with the initial condition* $x(0) = -1$.

***Exercise* 1.6.4:** *Consider* $x' = \sin x$.

a) *Draw the phase diagram for* $-4\pi \leq x \leq 4\pi$. *On this interval mark the critical points stable or unstable.*

b) *Sketch typical solutions of the equation.*

c) *Find* $\lim_{t\to\infty} x(t)$ *for the solution with the initial condition* $x(0) = 1$.

***Exercise* 1.6.5:** *Suppose* $f(x)$ *is positive for* $0 < x < 1$, *it is zero when* $x = 0$ *and* $x = 1$, *and it is negative for all other* x.

a) *Draw the phase diagram for* $x' = f(x)$, *find the critical points, and mark them stable or unstable.*

b) *Sketch typical solutions of the equation.*

c) *Find* $\lim_{t\to\infty} x(t)$ *for the solution with the initial condition* $x(0) = 0.5$.

***Exercise* 1.6.6:** *Start with the logistic equation* $\frac{dx}{dt} = kx(M-x)$. *Suppose we modify our harvesting. That is we will only harvest an amount proportional to current population. In other words, we harvest* hx *per unit of time for some* $h > 0$ *(similar to earlier example with* h *replaced with* hx*).*

a) *Construct the differential equation.*

b) *Show that if* $kM > h$, *then the equation is still logistic.*

c) *What happens when* $kM < h$?

***Exercise* 1.6.7:** *A disease is spreading through the country. Let* x *be the number of people infected. Let the constant* S *be the number of people susceptible to infection. The infection rate* $\frac{dx}{dt}$ *is proportional to the product of already infected people,* x, *and the number of susceptible but uninfected people,* $S - x$.

a) *Write down the differential equation.*

b) *Supposing* $x(0) > 0$, *that is, some people are infected at time* $t = 0$, *what is* $\lim_{t\to\infty} x(t)$.

c) *Does the solution to part b) agree with your intuition? Why or why not?*

Exercise* 1.6.101*: *Let* $x' = (x-1)(x-2)x^2$.

a) *Sketch the phase diagram and find critical points.*

b) *Classify the critical points.*

c) *If* $x(0) = 0.5$, *then find* $\lim_{t\to\infty} x(t)$.

Exercise* 1.6.102*: *Let* $x' = e^{-x}$.

a) *Find and classify all critical points.*

b) *Find* $\lim_{t\to\infty} x(t)$ *given any initial condition.*

Exercise* 1.6.103*: *Assume that a population of fish in a lake satisfies* $\frac{dx}{dt} = kx(M-x)$. *Now suppose that fish are continually added at* A *fish per unit of time.*

a) *Find the differential equation for* x.

b) *What is the new limiting population?*

Exercise* 1.6.104*: *Suppose* $\frac{dx}{dt} = (x-\alpha)(x-\beta)$ *for two numbers* $\alpha < \beta$.

a) *Find the critical points, and classify them.*

For b), c), d), find $\lim_{t\to\infty} x(t)$ *based on the phase diagram.*

b) $x(0) < \alpha$,

c) $\alpha < x(0) < \beta$,

d) $\beta < x(0)$.

1.7 Numerical methods: Euler's method

Note: 1 lecture, can safely be skipped, §2.4 in [EP], §8.1 in [BD]

Unless $f(x, y)$ is of a special form, it is generally very hard if not impossible to get a nice formula for the solution of the problem

$$y' = f(x, y), \qquad y(x_0) = y_0.$$

If the equation can be solved in closed form, we should do that. But what if we have an equation that cannot be solved in closed form? What if we want to find the value of the solution at some particular x? Or perhaps we want to produce a graph of the solution to inspect the behavior. In this section we will learn about the basics of numerical approximation of solutions.

The simplest method for approximating a solution is *Euler's method**. It works as follows: Take x_0 and y_0 and compute the slope $k = f(x_0, y_0)$. The slope is the change in y per unit change in x. Follow the line for an interval of length h on the x-axis. Hence if $y = y_0$ at x_0, then we say that y_1, the approximate value of y at $x_1 = x_0 + h$, is $y_1 = y_0 + hk$. Rinse, repeat! Let $k = f(x_1, y_1)$, and then compute $x_2 = x_1 + h$, and $y_2 = y_1 + hk$. Now compute x_3 and y_3 using x_2 and y_2, etc. Consider the equation $y' = y^2/3$, $y(0) = 1$, and $h = 1$. Then $x_0 = 0$ and $y_0 = 1$. We compute

$$x_1 = x_0 + h = 0 + 1 = 1, \qquad y_1 = y_0 + h\,f(x_0, y_0) = 1 + 1 \cdot 1/3 = 4/3 \approx 1.333,$$
$$x_2 = x_1 + h = 1 + 1 = 2, \qquad y_2 = y_1 + h\,f(x_1, y_1) = 4/3 + 1 \cdot \frac{(4/3)^2}{3} = 52/27 \approx 1.926.$$

We then draw an approximate graph of the solution by connecting the points (x_0, y_0), (x_1, y_1), (x_2, y_2),.... See Figure 1.16 on the following page for the first two steps of the method.

More abstractly, for any $i = 0, 1, 2, 3, \ldots$, we compute

$$x_{i+1} = x_i + h, \qquad y_{i+1} = y_i + h\,f(x_i, y_i).$$

The line segments we get are an approximate graph of the solution. Generally it is not exactly the solution. See Figure 1.17 on the next page for the plot of the real solution and the approximation.

We continue with the equation $y' = y^2/3$, $y(0) = 1$. Let us try to approximate $y(2)$ using Euler's method. In Figures 1.16 and 1.17 we have graphically approximated $y(2)$ with step size 1. With step size 1, we have $y(2) \approx 1.926$. The real answer is 3. We are approximately 1.074 off. Let us halve the step size. Computing y_4 with $h = 0.5$, we find that $y(2) \approx 2.209$, so an error of about 0.791. Table 1.1 on page 57 gives the values computed for various parameters.

---

*Named after the Swiss mathematician Leonhard Paul Euler (1707–1783). The correct pronunciation of the name sounds more like "oiler."

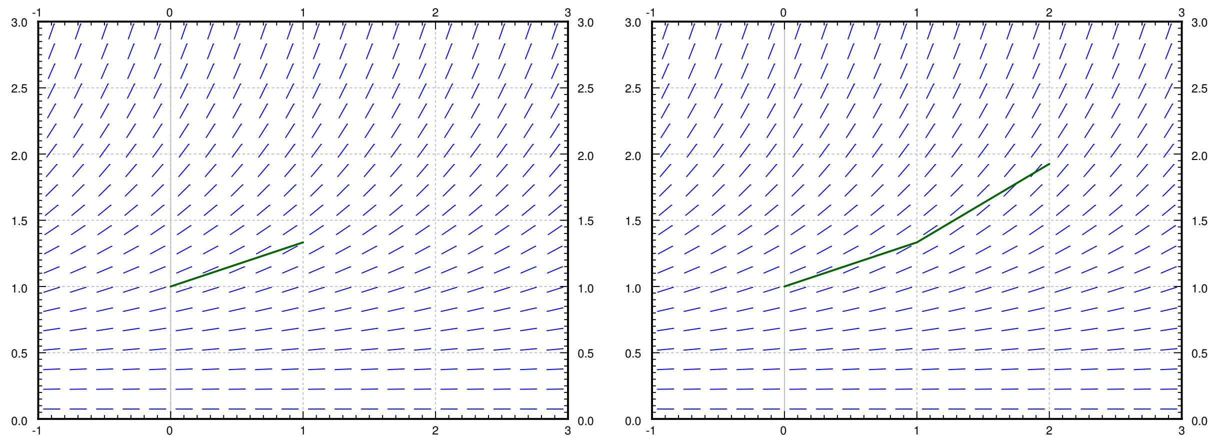


Figure 1.16: First two steps of Euler's method with $h = 1$ for the equation $y' = \frac{y^2}{3}$ with initial conditions $y(0) = 1$.

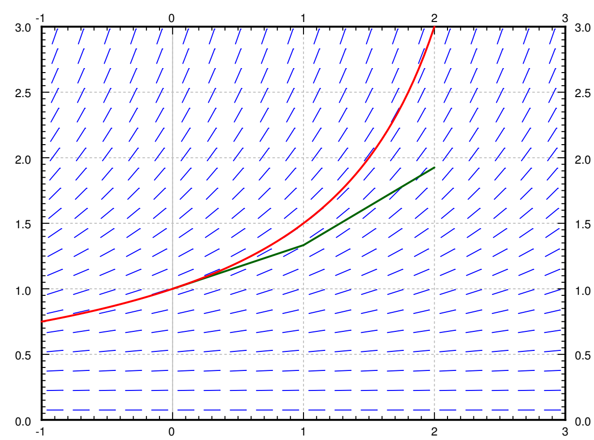


Figure 1.17: Two steps of Euler's method (step size 1) and the exact solution for the equation $y' = \frac{y^2}{3}$ with initial conditions $y(0) = 1$.

***Exercise* 1.7.1:** *Solve this equation exactly and show that $y(2) = 3$.*

The difference between the actual solution and the approximate solution is called the error. We usually talk about the size of the error and we do not care much about its sign.

$$\text{Error} = \left|\text{Actual } y - \text{Approximate } y\right|.$$

The point is, we do not know the real solution. If we knew the error exactly, we would know the actual solution . . . so what is the point of doing the approximation?

Note that except for the first few times, each time we halve the h, the error approximately halves. Halving of the error is a general feature of Euler's method as it is a *first order method*.

h	Approximate $y(2)$	Error	$\frac{\text{Error}}{\text{Previous error}}$
1	1.92593	1.07407	
0.5	2.20861	0.79139	0.73681
0.25	2.47250	0.52751	0.66656
0.125	2.68034	0.31966	0.60599
0.0625	2.82040	0.17960	0.56184
0.03125	2.90412	0.09588	0.53385
0.015625	2.95035	0.04965	0.51779
0.0078125	2.97472	0.02528	0.50913

Table 1.1: Euler's method approximation of $y(2)$ where of $y' = y^2/3$, $y(0) = 1$.

A simple improvement of the Euler method, see the exercises, produces a second order method. A second order method reduces the error to approximately one quarter every time we halve the interval. The order being "second" means the squaring in $1/4 = 1/2 \times 1/2 = (1/2)^2$.

To get the error to be within 0.1 of the answer, we had to do 64 steps. To get it to within 0.01, we would have to halve another three or four times, meaning doing 512 to 1024 steps. The improved Euler method from the exercises should quarter the error every time we halve the interval, so we would have to do (approximately) half as many "halvings" to get the same error. This reduction can be a big deal. With 10 halvings (starting at $h = 1$) we have 1024 steps, whereas with 5 halvings we only have to do 32 steps, assuming that the error was comparable to start with. A computer may not care about this difference for a problem this simple, but suppose each step would take a second to compute (the function may be substantially more difficult to compute than $y^2/3$). Then the difference is 32 seconds versus about 17 minutes. We are not being altogether fair, a second order method would probably double the time to do each step. Even so, it is 1 minute versus 17 minutes. Next, suppose that we have to repeat such a calculation for different parameters a thousand times. You get the idea.

In practice, we do not know how large the error is! How do we know what is the right step size? Well, essentially, we keep halving the interval, and if we are lucky, we can estimate the error from a few of these calculations and the assumption that the error goes down by a factor of one half each time (if we are using standard Euler).

***Exercise* 1.7.2:** *In the table above, suppose you do not know the error. Take the approximate values of the function in the last two lines, assume that the error goes down by a factor of 2. Can you estimate the error in the last time from this? Does it (approximately) agree with the table? Now do it for the first two rows. Does this agree with the table?*

Let us talk a little bit more about the example $y' = y^2/3$, $y(0) = 1$. Suppose that instead of $y(2)$ we wish to find $y(3)$. Table 1.2 on the next page lists the results of this effort for successive halvings of h. What is going on here? Well, you should solve the equation exactly and you will notice that the solution does not exist at $x = 3$. In fact, the solution goes to infinity when you approach $x = 3$.

h	Approximate $y(3)$
1	3.16232
0.5	4.54329
0.25	6.86079
0.125	10.80321
0.0625	17.59893
0.03125	29.46004
0.015625	50.40121
0.0078125	87.75769

Table 1.2: Attempts to use Euler's to approximate $y(3)$ where of $y' = y^2/3$, $y(0) = 1$.

Another case where things go bad is if the solution oscillates wildly near some point. The solution may exist at all points, but even a much better numerical method than Euler would need an insanely small step size to approximate the solution with reasonable precision. And computers might not be able to easily handle such a small step size.

In real applications we would not use a simple method such as Euler's. The simplest method that would probably be used in a real application is the standard Runge–Kutta method (see exercises). That is a fourth order method, meaning that if we halve the interval, the error generally goes down by a factor of 16 (it is fourth order as $1/16 = 1/2 \times 1/2 \times 1/2 \times 1/2$).

Choosing the right method to use and the right step size can be very tricky. There are several competing factors to consider.

- Computational time: Each step takes computer time. Even if the function f is simple to compute, we do it many times over. Large step size means faster computation, but perhaps not the right precision.
- Roundoff errors: Computers only compute with a certain number of significant digits. Errors introduced by rounding numbers off during our computations become noticeable when the step size becomes too small relative to the quantities we are working with. So reducing step size may in fact make errors worse. There is a certain optimum step size such that the precision increases as we approach it, but then starts getting worse as we make our step size smaller still. Trouble is: this optimum may be hard to find.
- Stability: Certain equations may be numerically unstable. What may happen is that the numbers never seem to stabilize no matter how many times we halve the interval.

We may need a ridiculously small interval size, which may not be practical due to roundoff errors or computational time considerations. Such problems are sometimes called *stiff*. In the worst case, the numerical computations might be giving us bogus numbers that look like a correct answer. Just because the numbers seem to have stabilized after successive halving, does not mean that we must have the right answer.

We have seen just the beginnings of the challenges that appear in real applications. Numerical approximation of solutions to differential equations is an active research area for engineers and mathematicians. For example, the general purpose method used for the ODE solver in Matlab and Octave (as of this writing) is a method that appeared in the literature only in the 1980s.

1.7.1 Exercises

***Exercise* 1.7.3:** *Consider* $\frac{dx}{dt} = (2t - x)^2$, $x(0) = 2$. *Use Euler's method with step size* $h = 0.5$ *to approximate* $x(1)$.

***Exercise* 1.7.4:** *Consider* $\frac{dx}{dt} = t - x$, $x(0) = 1$.

a) *Use Euler's method with step sizes* $h = 1, 1/2, 1/4, 1/8$ *to approximate* $x(1)$.
b) *Solve the equation exactly.*
c) *Describe what happens to the errors for each* h *you used. That is, find the factor by which the error changed each time you halved the interval.*

***Exercise* 1.7.5:** *Approximate the value of* e *by looking at the initial value problem* $y' = y$ *with* $y(0) = 1$ *and approximating* $y(1)$ *using Euler's method with a step size of* 0.2.

***Exercise* 1.7.6:** *Example of numerical instability: Take* $y' = -5y$, $y(0) = 1$. *We know that the solution should decay to zero as* x *grows. Using Euler's method, start with* $h = 1$ *and compute* y_1, y_2, y_3, y_4 *to try to approximate* $y(4)$. *What happened? Now halve the interval. Keep halving the interval and approximating* $y(4)$ *until the numbers you are getting start to stabilize (that is, until they start going towards zero). Note: You might want to use a calculator.*

The simplest method used in practice is the *Runge–Kutta method*. Consider $\frac{dy}{dx} = f(x, y)$, $y(x_0) = y_0$, and a step size h. Everything is the same as in Euler's method, except the computation of y_{i+1} and x_{i+1}.

$$\begin{aligned} k_1 &= f(x_i, y_i), & & \\ k_2 &= f\big(x_i + h/2, y_i + k_1(h/2)\big), & x_{i+1} &= x_i + h, \\ k_3 &= f\big(x_i + h/2, y_i + k_2(h/2)\big), & y_{i+1} &= y_i + \frac{k_1 + 2k_2 + 2k_3 + k_4}{6}\, h, \\ k_4 &= f(x_i + h, y_i + k_3 h). & & \end{aligned}$$

***Exercise* 1.7.7:** *Consider* $\frac{dy}{dx} = yx^2$, $y(0) = 1$.

a) Use Runge–Kutta (see above) with step sizes $h = 1$ *and* $h = 1/2$ *to approximate* $y(1)$.

b) Use Euler's method with $h = 1$ *and* $h = 1/2$.

c) Solve exactly, find the exact value of $y(1)$*, and compare.*

***Exercise* 1.7.101:** *Let* $x' = \sin(xt)$*, and* $x(0) = 1$*. Approximate* $x(1)$ *using Euler's method with step sizes 1, 0.5, 0.25. Use a calculator and compute up to 4 decimal digits.*

***Exercise* 1.7.102:** *Let* $x' = 2t$*, and* $x(0) = 0$.

a) Approximate $x(4)$ *using Euler's method with step sizes 4, 2, and 1.*

b) Solve exactly, and compute the errors.

c) Compute the factor by which the errors changed.

***Exercise* 1.7.103:** *Let* $x' = xe^{xt+1}$*, and* $x(0) = 0$.

a) Approximate $x(4)$ *using Euler's method with step sizes 4, 2, and 1.*

b) Guess an exact solution based on part a) and compute the errors.

There is a simple way to improve Euler's method to make it a second order method by doing just one extra step. Consider $\frac{dy}{dx} = f(x, y)$, $y(x_0) = y_0$, and a step size h. What we do is to pretend we compute the next step as in Euler, that is, we start with (x_i, y_i), we compute a slope $k_1 = f(x_i, y_i)$, and then look at the point $(x_i + h, y_i + k_1 h)$. Instead of letting our new point be $(x_i + h, y_i + k_1 h)$, we compute the slope at that point, call it k_2, and then take the average of k_1 and k_2, hoping that the average is going to be closer to the actual slope on the interval from x_i to $x_i + h$. And we are correct, if we halve the step, the error should go down by a factor of $2^2 = 4$. To summarize, the setup is the same as for regular Euler, except the computation of y_{i+1} and x_{i+1}.

$$\begin{aligned} k_1 &= f(x_i, y_i), & x_{i+1} &= x_i + h, \\ k_2 &= f(x_i + h, y_i + k_1 h), & y_{i+1} &= y_i + \frac{k_1 + k_2}{2}\, h. \end{aligned}$$

***Exercise* 1.7.104:** *Consider* $\frac{dy}{dx} = x + y$, $y(0) = 1$.

a) Use the improved Euler's method (see above) with step sizes $h = 1/4$ *and* $h = 1/8$ *to approximate* $y(1)$.

b) Use Euler's method with $h = 1/4$ *and* $h = 1/8$.

c) Solve exactly, find the exact value of $y(1)$.

d) Compute the errors, and the factors by which the errors changed.

1.8 Exact equations

Note: 1–2 lectures, can safely be skipped, §1.6 in [EP], §2.6 in [BD]

A type of equation that comes up quite often in physics and engineering is an *exact equation*. Suppose $F(x, y)$ is a function of two variables, which we call the *potential function*. The naming should suggest potential energy, or electric potential. Exact equations and potential functions appear when there is a conservation law at play, such as conservation of energy. Let us make up a simple example. Consider

$$F(x, y) = x^2 + y^2.$$

We are interested in the lines of constant energy, that is, lines where the energy is conserved: the curves where $F(x, y) = C$, for some constant C. In our example, the curves $x^2 + y^2 = C$ are circles. See Figure 1.18.

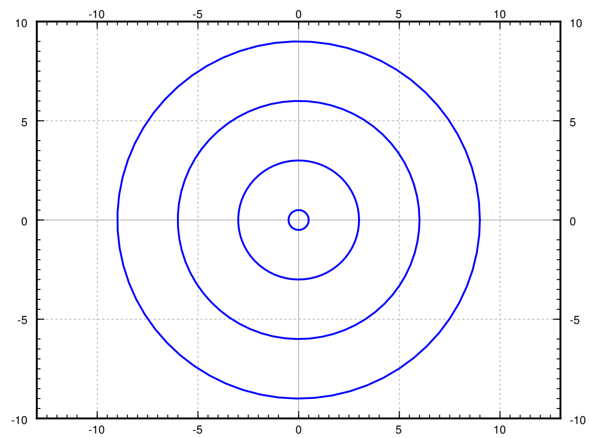


Figure 1.18: Solutions to $F(x, y) = x^2 + y^2 = C$ for various C.

We take the *total derivative* of F:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy.$$

For convenience, we will use the notation $F_x = \frac{\partial F}{\partial x}$ and $F_y = \frac{\partial F}{\partial y}$. In our example,

$$dF = 2x\, dx + 2y\, dy.$$

We apply the total derivative to $F(x, y) = C$, to find the differential equation $dF = 0$. The differential equation we obtain in such a way has the form

$$M\, dx + N\, dy = 0, \qquad \text{or} \qquad M + N\,\frac{dy}{dx} = 0.$$

An equation of this form is called *exact* if it was obtained as $dF = 0$ for some potential function F. In our simple example, we obtain the equation

$$2x\, dx + 2y\, dy = 0, \qquad \text{or} \qquad 2x + 2y\,\frac{dy}{dx} = 0.$$

Since we obtained this equation by differentiating $x^2 + y^2 = C$, the equation is exact. We often wish to solve for y in terms of x. In our example,

$$y = \pm\sqrt{C - x^2}.$$

An interpretation of the setup is that at each point (x, y), $\vec{v} = (M, N)$ is a vector in the plane, that is, a direction and a magnitude. As M and N are functions of (x, y), we have a *vector field*. The particular vector field $\vec{v}$ that comes from an exact equation is a so-called *conservative vector field*, that is, a vector field that comes with a potential function $F(x, y)$, such that

$$\vec{v} = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y} \right).$$

Let γ be a path in the plane starting at (x_1, y_1) and ending at (x_2, y_2). If we think of $\vec{v}$ as force, then the work required to move along γ is

$$\int_\gamma \vec{v}(\vec{r}) \cdot d\vec{r} = \int_\gamma M\, dx + N\, dy = F(x_2, y_2) - F(x_1, y_1).$$

In other words, the work done only depends on endpoints, that is, where we start and where we end. For example, suppose F is gravitational potential. The derivative of F given by $\vec{v}$ is the gravitational force. What we are saying is that the work required to move a heavy box from the ground floor to the roof, only depends on the change in potential energy. That is, the work done is the same no matter what path we took; if we took the stairs or the elevator. Although if we took the elevator, the elevator is doing the work for us. The curves $F(x, y) = C$ are those where no work need be done, such as the heavy box sliding along without accelerating or breaking on a perfectly flat roof, on a cart with incredibly well oiled wheels.

An exact equation is a conservative vector field, and the implicit solution of this equation is the potential function.

1.8.1 Solving exact equations

Now you, the reader, should ask: Where did we solve a differential equation? Well, in applications we generally know M and N, but we do not know F. That is, we may have just started with $2x + 2y\frac{dy}{dx} = 0$, or perhaps even

$$x + y\frac{dy}{dx} = 0.$$

It is up to us to find some potential F that works. Many different F will work; adding a constant to F does not change the equation. Once we have a potential function F, the equation $F\big(x, y(x)\big) = C$ gives an implicit solution of the ODE.

Example 1.8.1: Let us find the general solution to $2x + 2y\frac{dy}{dx} = 0$. Forget we knew what F was.

If we know that this is an exact equation, we start looking for a potential function F. We have $M = 2x$ and $N = 2y$. If F exists, it must be such that $F_x(x, y) = 2x$. Integrate in the x variable to find

$$F(x, y) = x^2 + A(y), \tag{1.5}$$

for some function $A(y)$. The function A is the "constant of integration," though it is only constant as far as x is concerned, and may still depend on y. Now differentiate (1.5) in y

and set it equal to N, which is what F_y is supposed to be:

$$2y = F_y(x, y) = A'(y).$$

Integrating, we find $A(y) = y^2$. We could add a constant of integration if we wanted to, but there is no need. We found $F(x, y) = x^2 + y^2$. Next for a constant C, we solve

$$F\big(x, y(x)\big) = C.$$

for y in terms of x. In this case, we obtain $y = \pm\sqrt{C - x^2}$ as we did before.

Exercise* 1.8.1*: *Why did we not need to add a constant of integration when integrating $A'(y) = 2y$? Add a constant of integration, say 3, and see what F you get. What is the difference from what we got above, and why does it not matter?*

The procedure, once we know that the equation is exact, is:

(i) Integrate $F_x = M$ in x resulting in $F(x, y) = \text{something} + A(y)$.

(ii) Differentiate this F in y, and set that equal to N, so that we may find $A(y)$ by integration.

The procedure can also be done by first integrating in y and then differentiating in x. Pretty easy huh? Let's try this again.

Example 1.8.2: Consider now $2x + y + xy\frac{dy}{dx} = 0$.

OK, so $M = 2x + y$ and $N = xy$. We try to proceed as before. Suppose F exists. Then $F_x(x, y) = 2x + y$. We integrate:

$$F(x, y) = x^2 + xy + A(y)$$

for some function $A(y)$. Differentiate in y and set equal to N:

$$N = xy = F_y(x, y) = x + A'(y).$$

But there is no way to satisfy this requirement! The function xy cannot be written as x plus a function of y. The equation is not exact; no potential function F exists.

Is there an easier way to check for the existence of F, other than failing in trying to find it? Turns out there is. Suppose $M = F_x$ and $N = F_y$. Then as long as the second derivatives are continuous,

$$\frac{\partial M}{\partial y} = \frac{\partial^2 F}{\partial y \partial x} = \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial N}{\partial x}.$$

Let us state it as a theorem. Usually this is called the Poincaré Lemma*.

Theorem 1.8.1 (Poincaré)**.** *If M and N are continuously differentiable functions of (x, y), and $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, then near any point there is a function $F(x, y)$ such that $M = \frac{\partial F}{\partial x}$ and $N = \frac{\partial F}{\partial y}$.*

*Named for the French polymath Jules Henri Poincaré (1854–1912).

The theorem doesn't give us a global F defined everywhere in the plane. In general, we can only find the potential locally, near some initial point. By this time, we have come to expect this from differential equations.

Let us return to Example 1.8.2, where $M = 2x + y$ and $N = xy$. Notice $M_y = 1$ and $N_x = y$, which are clearly not equal. The equation is not exact.

Example 1.8.3: Solve

$$\frac{dy}{dx} = \frac{-2x - y}{x - 1}, \qquad y(0) = 1.$$

We write the equation as

$$(2x + y) + (x - 1)\frac{dy}{dx} = 0,$$

so $M = 2x + y$ and $N = x - 1$. Then

$$M_y = 1 = N_x.$$

The equation is exact. Integrating M in x, we find

$$F(x, y) = x^2 + xy + A(y).$$

Differentiating in y and setting to N, we find

$$x - 1 = x + A'(y).$$

So $A'(y) = -1$, and $A(y) = -y$ will work. We obtain $F(x, y) = x^2 + xy - y$, so the implicit solution is $x^2 + xy - y = C$. First we find C. As $y(0) = 1$, we have $F(0, 1) = C$. Therefore, $0^2 + 0 \times 1 - 1 = C$, so $C = -1$. Now we solve $x^2 + xy - y = -1$ for y to get

$$y = \frac{-x^2 - 1}{x - 1}.$$

Example 1.8.4: Solve

$$\frac{-y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy = 0, \qquad y(1) = 2.$$

We leave to the reader to check that $M_y = N_x$.

This vector field (M, N) is not conservative if considered as a vector field of the entire plane minus the origin. The problem is that if the curve γ is a circle around the origin, say starting at $(1, 0)$ and ending at $(1, 0)$ going counterclockwise, then if F existed we would expect

$$0 = F(1, 0) - F(1, 0) = \int_\gamma F_x \, dx + F_y \, dy = \int_\gamma \frac{-y}{x^2 + y^2} \, dx + \frac{x}{x^2 + y^2} \, dy = 2\pi.$$

That is nonsense! We leave the computation of the path integral to the interested reader, or you can consult your multivariable calculus textbook. So there is no potential function F defined everywhere outside the origin $(0, 0)$.

If we think back to the theorem, it does not guarantee such a function anyway. It only guarantees a potential function locally, that is, only in some region near the initial point. As $y(1) = 2$, we start at the point $(1, 2)$. Considering $x > 0$ and integrating M in x or N in y, we find

$$F(x, y) = \arctan\left(y/x\right).$$

The implicit solution is $\arctan\left(y/x\right) = C$. Solving, $y = \tan(C)x$. That is, the solution is a straight line. Solving $y(1) = 2$ gives us that $\tan(C) = 2$, and so $y = 2x$ is the desired solution. See Figure 1.19, and note that the solution only exists for $x > 0$.

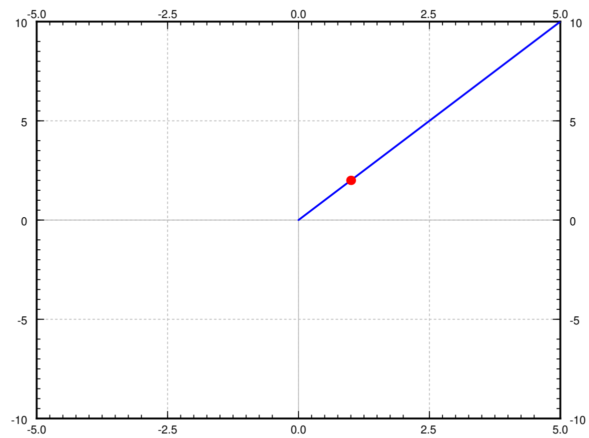


Figure 1.19: Solution to $\frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy = 0$, $y(1) = 2$, *with initial point marked.*

Example 1.8.5: Solve

$$x^2 + y^2 + 2y(x + 1)\frac{dy}{dx} = 0.$$

The reader should check that this equation is exact. Let $M = x^2 + y^2$ and $N = 2y(x + 1)$. We follow the procedure for exact equations

$$F(x, y) = \frac{1}{3}x^3 + xy^2 + A(y),$$

and

$$2y(x + 1) = 2xy + A'(y).$$

Therefore $A'(y) = 2y$ or $A(y) = y^2$ and $F(x, y) = \frac{1}{3}x^3 + xy^2 + y^2$. We try to solve $F(x, y) = C$. We easily solve for y^2 and then just take the square root:

$$y^2 = \frac{C - (1/3)x^3}{x + 1}, \qquad \text{so} \qquad y = \pm\sqrt{\frac{C - (1/3)x^3}{x + 1}}.$$

When $x = -1$, the term in front of $\frac{dy}{dx}$ is zero, and our explicit solution is not valid. The given equation has no solution (for $y(x)$) near $x = -1$, but the equation $(x^2+y^2)\,dx+2y(x+1)\,dy = 0$ does have a solution $x = -1$. In fact, one could solve for x in terms of y for any initial condition. The solution is messy, so we leave it as $\frac{1}{3}x^3 + xy^2 + y^2 = C$.

1.8.2 Integrating factors

Sometimes an equation $M\,dx + N\,dy = 0$ is not exact, but it can be made exact by multiplying with a function $u(x, y)$. That is, perhaps for some nonzero function $u(x, y)$,

$$u(x, y)M(x, y)\,dx + u(x, y)N(x, y)\,dy = 0$$

is exact. Any solution to this new equation is also a solution to $M\,dx + N\,dy = 0$.

In fact, a linear equation

$$\frac{dy}{dx} + p(x)y = f(x), \qquad \text{or} \qquad \bigl(p(x)y - f(x)\bigr)\,dx + dy = 0$$

is always such an equation. Let $r(x) = e^{\int p(x)\,dx}$ be the integrating factor for a linear equation. Multiply the equation by $r(x)$ and write it in the form of $M + N\frac{dy}{dx} = 0$.

$$r(x)p(x)y - r(x)f(x) + r(x)\frac{dy}{dx} = 0.$$

Then $M = r(x)p(x)y - r(x)f(x)$, so $M_y = r(x)p(x)$, while $N = r(x)$, so $N_x = r'(x) = r(x)p(x)$. In other words, we have an exact equation. Integrating factors for linear functions are just a special case of integrating factors for exact equations.

But how do we find the integrating factor u? Well, given an equation

$$M\,dx + N\,dy = 0,$$

u should be a function such that

$$\frac{\partial}{\partial y}\bigl[uM\bigr] = u_y M + u M_y = \frac{\partial}{\partial x}\bigl[uN\bigr] = u_x N + u N_x.$$

Therefore,

$$(M_y - N_x)u = u_x N - u_y M.$$

At first it may seem we replaced one differential equation by another. True, but all hope is not lost.

A strategy that often works is to look for a u that is a function of x alone, or a function of y alone. If u is a function of x alone, that is $u(x)$, then we write $u'(x)$ instead of u_x, and u_y is just zero. Then

$$\frac{M_y - N_x}{N}u = u'.$$

In particular, $\frac{M_y - N_x}{N}$ ought to be a function of x alone (not depend on y). If so, then we have a linear equation

$$u' - \frac{M_y - N_x}{N}u = 0.$$

Letting $P(x) = \frac{M_y - N_x}{N}$, we solve using the standard integrating factor method, to find $u(x) = Ce^{\int P(x)\,dx}$. The constant in the solution is not relevant, we need any nonzero solution, so we take $C = 1$. Then $u(x) = e^{\int P(x)\,dx}$ is the integrating factor.

Similarly, we could try a function of the form $u(y)$. Then

$$\frac{M_y - N_x}{M} u = -u'.$$

In particular, $\frac{M_y - N_x}{M}$ ought to be a function of y alone. If so, we have a linear equation

$$u' + \frac{M_y - N_x}{M} u = 0.$$

Letting $Q(y) = \frac{M_y - N_x}{M}$, we find $u(y) = Ce^{-\int Q(y)\,dy}$. We take $C = 1$. So $u(y) = e^{-\int Q(y)\,dy}$ is the integrating factor.

Example 1.8.6: Solve

$$\frac{x^2 + y^2}{x + 1} + 2y\frac{dy}{dx} = 0.$$

Let $M = \frac{x^2+y^2}{x+1}$ and $N = 2y$. Compute

$$M_y - N_x = \frac{2y}{x+1} - 0 = \frac{2y}{x+1}.$$

As this is not zero, the equation is not exact. We notice

$$P(x) = \frac{M_y - N_x}{N} = \frac{2y}{x+1}\frac{1}{2y} = \frac{1}{x+1}$$

is a function of x alone. We compute the integrating factor

$$e^{\int P(x)\,dx} = e^{\ln(x+1)} = x + 1.$$

We multiply our given equation by $(x + 1)$ to obtain

$$x^2 + y^2 + 2y(x + 1)\frac{dy}{dx} = 0,$$

which is an exact equation that we solved in Example 1.8.5. The solution was

$$y = \pm\sqrt{\frac{C - (1/3)x^3}{x + 1}}.$$

Example 1.8.7: Solve

$$y^2 + (xy + 1)\frac{dy}{dx} = 0.$$

First compute

$$M_y - N_x = 2y - y = y.$$

As this is not zero, the equation is not exact. We observe

$$Q(y) = \frac{M_y - N_x}{M} = \frac{y}{y^2} = \frac{1}{y}$$

is a function of y alone. We compute the integrating factor

$$e^{-\int Q(y)\,dy} = e^{-\ln y} = \frac{1}{y}.$$

Therefore, we look at the exact equation

$$y + \frac{xy+1}{y}\frac{dy}{dx} = 0.$$

The reader should double check that this equation is exact. We follow the procedure for exact equations

$$F(x,y) = xy + A(y),$$

and

$$\frac{xy+1}{y} = x + \frac{1}{y} = x + A'(y). \tag{1.6}$$

Consequently, $A'(y) = 1/y$ or $A(y) = \ln|y|$. Thus $F(x,y) = xy + \ln|y|$. It is not possible to solve $F(x,y) = C$ for y in terms of elementary functions, so let us be content with the implicit solution:

$$xy + \ln|y| = C.$$

We are looking for the general solution and we divided by y above. We should check what happens when $y = 0$, as the equation itself makes perfect sense in that case. We plug in $y = 0$ to find the equation is satisfied. So $y = 0$ is also a solution.

1.8.3 Exercises

***Exercise* 1.8.2:** *Solve the following exact equations, implicit general solutions will suffice:*

a) $(2xy + x^2)\,dx + (x^2 + y^2 + 1)\,dy = 0$

b) $x^5 + y^5\frac{dy}{dx} = 0$

c) $e^x + y^3 + 3xy^2\frac{dy}{dx} = 0$

d) $(x+y)\cos(x) + \sin(x) + \sin(x)y' = 0$

***Exercise* 1.8.3:** *Find the integrating factor for the following equations making them into exact equations:*

a) $e^{xy}\,dx + \frac{y}{x}e^{xy}\,dy = 0$

b) $\frac{e^x + y^3}{y^2}\,dx + 3x\,dy = 0$

c) $4(y^2 + x)\,dx + \frac{2x + 2y^2}{y}\,dy = 0$

d) $2\sin(y)\,dx + x\cos(y)\,dy = 0$

***Exercise* 1.8.4:** *Suppose you have an equation of the form:* $f(x) + g(y)\frac{dy}{dx} = 0$.

a) *Show it is exact.*

b) *Find the form of the potential function in terms of* f *and* g.

***Exercise* 1.8.5:** *Suppose that we have the equation* $f(x)\,dx - dy = 0$.

a) *Is this equation exact?*

b) *Find the general solution using a definite integral.*

***Exercise* 1.8.6:** *Find the potential function* $F(x, y)$ *of the exact equation* $\frac{1+xy}{x}\,dx + \left(1/y + x\right)dy = 0$ *in two different ways.*

a) *Integrate* M *in terms of* x *and then differentiate in* y *and set to* N.

b) *Integrate* N *in terms of* y *and then differentiate in* x *and set to* M.

***Exercise* 1.8.7:** *A function* $u(x, y)$ *is said to be a* harmonic function *if* $u_{xx} + u_{yy} = 0$.

a) *Show if* u *is harmonic,* $-u_y\,dx + u_x\,dy = 0$ *is an exact equation. So there exists (at least locally) the so-called* harmonic conjugate *function* $v(x, y)$ *such that* $v_x = -u_y$ *and* $v_y = u_x$.

Verify that the following u *are harmonic and find the corresponding harmonic conjugates* v:

b) $u = 2xy$ *c)* $u = e^x \cos y$ *d)* $u = x^3 - 3xy^2$

***Exercise* 1.8.101:** *Solve the following exact equations, implicit general solutions will suffice:*

a) $\cos(x) + ye^{xy} + xe^{xy}y' = 0$ *b)* $(2x + y)\,dx + (x - 4y)\,dy = 0$

c) $e^x + e^y\frac{dy}{dx} = 0$ *d)* $(3x^2 + 3y)\,dx + (3y^2 + 3x)\,dy = 0$

***Exercise* 1.8.102:** *Find the integrating factor for the following equations making them into exact equations:*

a) $\frac{1}{y}\,dx + 3y\,dy = 0$ *b)* $dx - e^{-x-y}\,dy = 0$

c) $\left(\frac{\cos(x)}{y^2} + \frac{1}{y}\right)dx + \frac{x}{y^2}\,dy = 0$ *d)* $\left(2y + \frac{y^2}{x}\right)dx + (2y + x)\,dy = 0$

***Exercise* 1.8.103:**

a) *Show that every separable equation* $y' = f(x)g(y)$ *can be written as an exact equation, and verify that it is indeed exact.*

b) *Rewrite* $y' = xy$ *as an exact equation, solve it, and verify that the solution is the same as it was in* *Example 1.3.1*.

Chapter 2

Systems of ODEs

2.1 Introduction to systems of ODEs

Note: 1 to 1.5 lectures, §4.1 in [EP], §7.1 in [BD]

2.1.1 Systems

Often we do not have just one dependent variable and one equation. And as we will see, we may end up with systems of several equations and several dependent variables even if we start with a single equation.

If we have several dependent variables, suppose $y_1, y_2, \ldots, y_n$, then we can have a differential equation involving all of them and their derivatives with respect to one independent variable x. For example, $y_1'' = f(y_1', y_2', y_1, y_2, x)$. Usually, when we have two dependent variables we have two equations such as

$$
\begin{aligned}
y_1'' &= f_1(y_1', y_2', y_1, y_2, x), \\
y_2'' &= f_2(y_1', y_2', y_1, y_2, x),
\end{aligned}
$$

for some functions f_1 and f_2. We call the above a *system of differential equations*. More precisely, the above is a *second order system* of ODEs as second order derivatives appear. The system

$$
\begin{aligned}
x_1' &= g_1(x_1, x_2, x_3, t), \\
x_2' &= g_2(x_1, x_2, x_3, t), \\
x_3' &= g_3(x_1, x_2, x_3, t),
\end{aligned}
$$

is a *first order system*, where x_1, x_2, x_3 are the dependent variables, and t is the independent variable.

The terminology for systems is essentially the same as for single equations. For the system above, a *solution* is a set of three functions $x_1(t)$, $x_2(t)$, $x_3(t)$, such that

$$\begin{aligned} x_1'(t) &= g_1\big(x_1(t), x_2(t), x_3(t), t\big), \\ x_2'(t) &= g_2\big(x_1(t), x_2(t), x_3(t), t\big), \\ x_3'(t) &= g_3\big(x_1(t), x_2(t), x_3(t), t\big). \end{aligned}$$

We usually also have an *initial condition*. Just like for single equations, we specify x_1, x_2, and x_3 for some fixed t. For example, $x_1(0) = a_1$, $x_2(0) = a_2$, $x_3(0) = a_3$, where a_1, a_2, and a_3 are some constants. For the second order system, we would also specify the first derivatives at a point. If we find a solution with some arbitrary constants in it, where by solving for the constants we find a solution for any initial condition, we call this solution the *general solution*. Best to look at a simple example.

Example 2.1.1: Sometimes a system is easy to solve by solving for one variable and then for the second variable. Take the first order system

$$\begin{aligned} y_1' &= y_1, \\ y_2' &= y_1 - y_2, \end{aligned}$$

with y_1, y_2 as the dependent variables and x as the independent variable. And consider initial conditions $y_1(0) = 1$, $y_2(0) = 2$.

We note that $y_1 = C_1 e^x$ is the general solution of the first equation. We then plug this y_1 into the second equation and get the equation $y_2' = C_1 e^x - y_2$, which is a linear first order equation that is easily solved for y_2. By the method of integrating factor we get

$$e^x y_2 = \frac{C_1}{2} e^{2x} + C_2,$$

or $y_2 = \frac{C_1}{2} e^x + C_2 e^{-x}$. The general solution to the system is, therefore,

$$y_1 = C_1 e^x, \qquad y_2 = \frac{C_1}{2} e^x + C_2 e^{-x}.$$

We solve for C_1 and C_2 given the initial conditions. We substitute $x = 0$ and find that $C_1 = 1$ and $C_2 = 3/2$. Thus the solution is $y_1 = e^x$, and $y_2 = (1/2) e^x + (3/2) e^{-x}$.

Generally, we will not be so lucky to be able to solve for each variable separately as in the example above, and we will have to solve for all variables at once. While we won't generally be able to solve for one variable and then the next, we will try to salvage as much as possible from this technique. It will turn out that in a certain sense we will still (try to) solve a bunch of single equations and put their solutions together. Let us not worry right now about how to solve systems yet.

We will mostly consider the *linear systems*. The example above is a so-called *linear first order system*. It is linear as none of the dependent variables or their derivatives appear in nonlinear functions or with powers higher than one (x, y, x' and y', constants, and functions of t can appear, but not xy or $(y')^2$ or x^3). Another, more complicated, example of a linear system is

$$\begin{aligned} y_1'' &= e^t y_1' + t^2 y_1 + 5 y_2 + \sin(t), \\ y_2'' &= t y_1' - y_2' + 2 y_1 + \cos(t). \end{aligned}$$

2.1.2 Applications

Let us consider some simple applications of systems and how to set up the equations.

Example 2.1.2: First, we consider salt and brine tanks, but this time water flows from one to the other and back. We again consider that the tanks are evenly mixed.

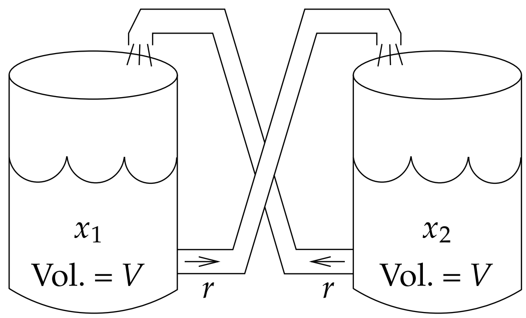


Figure 2.1: A closed system of two brine tanks.

Suppose we have two tanks, each containing volume V liters of salt brine. The amount of salt in the first tank is x_1 grams, and the amount of salt in the second tank is x_2 grams. The liquid is perfectly mixed and flows at the rate r liters per second out of each tank into the other. See Figure 2.1.

The rate of change of x_1, that is x_1', is the rate of salt coming in minus the rate going out. The rate coming in is the density of the salt in tank 2, that is $\frac{x_2}{V}$, times the rate r. The rate coming out is the density of the salt in tank 1, that is $\frac{x_1}{V}$, times the rate r. In other words it is

$$x_1' = \frac{x_2}{V} r - \frac{x_1}{V} r = \frac{r}{V} x_2 - \frac{r}{V} x_1 = \frac{r}{V}(x_2 - x_1).$$

Similarly we find the rate x_2', where the roles of x_1 and x_2 are reversed. All in all, the system of ODEs for this problem is

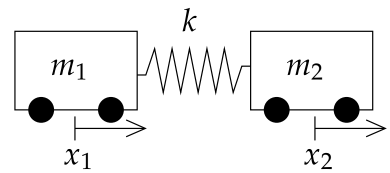
$$\begin{aligned} x_1' &= \frac{r}{V}(x_2 - x_1), \\ x_2' &= \frac{r}{V}(x_1 - x_2). \end{aligned}$$

In this system we cannot solve for x_1 or x_2 separately. We must solve for both x_1 and x_2 at once, which is intuitively clear since the amount of salt in one tank affects the amount in the other. We can't know x_1 before we know x_2, and vice versa.

We do not yet know how to find all the solutions, but intuitively we can at least find some solutions. Suppose we know that initially the tanks have the same amount of salt. That is, we have an initial condition such as $x_1(0) = x_2(0) = C$. Then clearly the amount of salt coming and out of each tank is the same, so the amounts are not changing. In other words, $x_1 = C$ and $x_2 = C$ (the constant functions) is a solution: $x_1' = x_2' = 0$, and $x_2 - x_1 = x_1 - x_2 = 0$, so the equations are satisfied.

Let us think about the setup a little bit more without solving it. Suppose the initial conditions are $x_1(0) = A$ and $x_2(0) = B$, for two different constants A and B. Since no salt is coming in or out of this closed system, the total amount of salt is constant. That is, $x_1 + x_2$ is constant, and so it equals $A + B$. Intuitively if A is bigger than B, then more salt will flow out of tank one than into it. Eventually, after a long time we would then expect the amount of salt in each tank to equalize. In other words, the solutions of both x_1 and x_2 should tend towards $\frac{A+B}{2}$. Once you know how to solve systems you will find out that this really is so.

Example 2.1.3: Let us look at a second order example. We return to the mass and spring setup, but this time we consider two masses.



Consider one spring with constant k and two masses m_1 and m_2. Think of the masses as carts that ride along a straight track with no friction. Let x_1 be the displacement of the first cart and x_2 be the displacement of the second cart. That is, we put the two carts somewhere with no tension on the spring, and we mark the position of the first and second cart and call those the zero positions. Then x_1 measures how far the first cart is from its zero position, and x_2 measures how far the second cart is from its zero position. The force exerted by the spring on the first cart is $k(x_2 - x_1)$, since $x_2 - x_1$ is how far the string is stretched (or compressed) from the rest position. The force exerted on the second cart is the opposite, thus the same thing with a negative sign. Newton's second law states that force equals mass times acceleration. So the system of equations is

$$\begin{aligned} m_1 x_1'' &= k(x_2 - x_1), \\ m_2 x_2'' &= -k(x_2 - x_1). \end{aligned}$$

Again, we cannot solve for the x_1 or x_2 variable separately. That we must solve for both x_1 and x_2 at once is intuitively clear, since where the first cart goes depends on exactly where the second cart goes and vice versa.

2.1.3 Changing to first order

Before we talk about how to handle systems, let us note that in some sense we need only consider first order systems. Let us take an n^{th} order differential equation

$$y^{(n)} = F(y^{(n-1)}, \ldots, y', y, x).$$

We define new variables $u_1, u_2, \ldots, u_n$ and write the system

$$\begin{aligned} u_1' &= u_2, \\ u_2' &= u_3, \\ &\vdots \\ u_{n-1}' &= u_n, \\ u_n' &= F(u_n, u_{n-1}, \ldots, u_2, u_1, x). \end{aligned}$$

We solve this system for $u_1, u_2, \ldots, u_n$. Once we have solved for the u, we can discard u_2 through u_n and let $y = u_1$. This y solves the original equation.

Example 2.1.4: Take $x''' = 2x'' + 8x' + x + t$. Letting $u_1 = x$, $u_2 = x'$, $u_3 = x''$, we find the system:

$$u_1' = u_2, \qquad u_2' = u_3, \qquad u_3' = 2u_3 + 8u_2 + u_1 + t.$$

A similar process can be followed for a system of higher order differential equations. For example, a system of k differential equations in k unknowns, all of order n, can be transformed into a first order system of $n \times k$ equations and $n \times k$ unknowns.

Example 2.1.5: Consider the system from the carts example,

$$m_1 x_1'' = k(x_2 - x_1), \qquad m_2 x_2'' = -k(x_2 - x_1).$$

Let $u_1 = x_1$, $u_2 = x_1'$, $u_3 = x_2$, $u_4 = x_2'$. The second order system becomes the first order system

$$u_1' = u_2, \qquad m_1 u_2' = k(u_3 - u_1), \qquad u_3' = u_4, \qquad m_2 u_4' = -k(u_3 - u_1).$$

Example 2.1.6: The idea works in reverse as well. Consider the system

$$x' = 2y - x, \qquad y' = x,$$

where the independent variable is t. We wish to solve for the initial conditions $x(0) = 1$, $y(0) = 0$.

If we differentiate the second equation, we get $y'' = x'$. We know what x' is in terms of x and y, and we know that $x = y'$. So,

$$y'' = x' = 2y - x = 2y - y'.$$

We now have the equation $y'' + y' - 2y = 0$. We know how to solve this equation and we find that $y = C_1 e^{-2t} + C_2 e^t$. Once we have y, we use the equation $y' = x$ to get x.

$$x = y' = -2C_1 e^{-2t} + C_2 e^t.$$

We solve for the initial conditions $1 = x(0) = -2C_1 + C_2$ and $0 = y(0) = C_1 + C_2$. Hence, $C_1 = -C_2$ and $1 = 3C_2$. So $C_1 = {-1}/3$ and $C_2 = 1/3$. Our solution is

$$x = \frac{2e^{-2t} + e^t}{3}, \qquad y = \frac{-e^{-2t} + e^t}{3}.$$

Exercise* 2.1.1*: *Plug in and check that this really is the solution.*

It is useful to go back and forth between systems and higher order equations for other reasons. For example, software for solving ODE numerically (approximation) is generally for first order systems. To use it, you take whatever ODE you want to solve and convert it to a first order system. It is not very hard to adapt computer code for the Euler or Runge–Kutta method for first order equations to handle first order systems. We simply treat the dependent variable not as a number but as a vector. In many mathematical computer languages there is almost no distinction in syntax.

2.1.4 Autonomous systems and vector fields

A system where the equations do not depend on the independent variable is called an *autonomous system*. For example the system $y' = 2y - x$, $y' = x$ is autonomous as t is the independent variable but does not appear in the equations.

For autonomous systems we can draw the so-called *direction field* or *vector field*, a plot similar to a slope field, but instead of giving a slope at each point, we give a direction (and a magnitude). The previous example, $x' = 2y - x$, $y' = x$, says that at the point (x, y) the direction in which we should travel to satisfy the equations should be the direction of the vector $(2y - x, x)$ with the speed equal to the magnitude of this vector. So we draw the vector $(2y - x, x)$ at the point (x, y) and we do this for many points on the xy-plane. For example, at the point $(1, 2)$ we draw the vector $\big(2(2) - 1, 1\big) = (3, 1)$, a vector pointing to the right and a little bit up, while at the point $(2, 1)$ we draw the vector $\big(2(1) - 2, 2\big) = (0, 2)$ a vector that points straight up. When drawing the vectors, we will scale down their size to fit many of them on the same direction field. If we drew the arrows at the actual size, the diagram would be a jumbled mess once you would draw more than a couple of arrows. So we scale them all so that not even the longest one interferes with the others. We are mostly interested in their direction and relative size. See Figure 2.2 on the next page.

We can draw a path of the solution in the plane. Suppose the solution is given by $x = f(t)$, $y = g(t)$. We pick an interval of t (say $0 \leq t \leq 2$ for our example) and plot all the points $\big(f(t), g(t)\big)$ for t in the selected range. The resulting picture is called the *phase portrait* (or phase plane portrait). The particular curve obtained is called the *trajectory* or *solution curve*. See an example plot in Figure 2.3 on the facing page. In the figure the solution starts at $(1, 0)$ and travels along the vector field for a distance of 2 units of t. We solved this system precisely, so we compute $x(2)$ and $y(2)$ to find $x(2) \approx 2.475$ and $y(2) \approx 2.457$. This point corresponds to the top right end of the plotted solution curve in the figure.

Notice the similarity to the diagrams we drew for autonomous systems in one dimension. But note how much more complicated things become when we allow just one extra dimension.

We can draw phase portraits and trajectories in the xy-plane even if the system is not autonomous. In this case, however, we cannot draw the direction field, since the field changes as t changes. For each t we would get a different direction field.

2.1.5 Picard's theorem

Perhaps before going further, let us mention that Picard's theorem on existence and uniqueness still holds for systems of ODE. Let us restate this theorem in the setting of

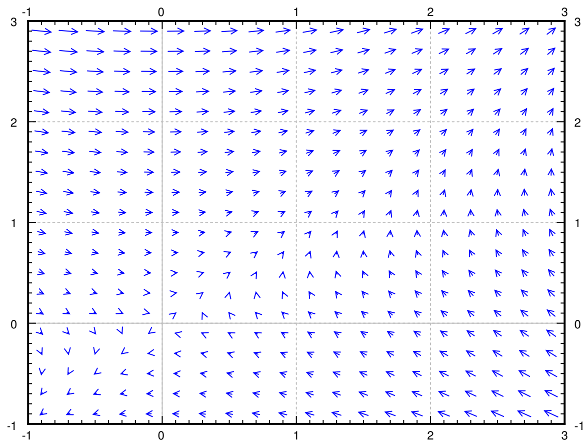


Figure 2.2: The direction field for $x' = 2y - x$, $y' = x$.

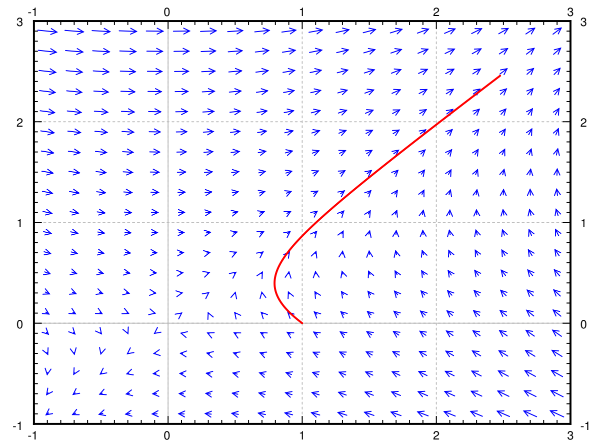


Figure 2.3: The direction field for $x' = 2y - x$, $y' = x$ with the trajectory of the solution starting at $(1, 0)$ for $0 \leq t \leq 2$.

systems. A general first order system is of the form

$$
\begin{aligned}
x_1' &= F_1(x_1, x_2, \ldots, x_n, t), \\
x_2' &= F_2(x_1, x_2, \ldots, x_n, t), \\
&\vdots \\
x_n' &= F_n(x_1, x_2, \ldots, x_n, t).
\end{aligned}
\tag{2.1}
$$

Theorem 2.1.1 (Picard's theorem on existence and uniqueness for systems)**.** *If for every $j = 1, 2, \ldots, n$ and every $k = 1, 2, \ldots, n$ each F_j is continuous and the derivative $\frac{\partial F_j}{\partial x_k}$ exists and is continuous near some $(x_1^0, x_2^0, \ldots, x_n^0, t^0)$, then a solution to (2.1) subject to the initial condition $x_1(t^0) = x_1^0$, $x_2(t^0) = x_2^0$, $\ldots$, $x_n(t^0) = x_n^0$ exists (at least for t in some small interval) and is unique.*

That is, a unique solution exists for any initial condition given that the system is reasonable (F_j and its partial derivatives in the x variables are continuous). As for single equations we may not have a solution for all time t, but at least for some short period of time.

As we can change any nth order ODE into a first order system, then we notice that this theorem provides also the existence and uniqueness of solutions for higher order equations that we have until now not stated explicitly.

2.1.6 Exercises

***Exercise* 2.1.2:** *Find the general solution of $x_1' = x_2 - x_1 + t$, $x_2' = x_2$.*

***Exercise* 2.1.3:** *Find the general solution of $x_1' = 3x_1 - x_2 + e^t$, $x_2' = x_1$.*

Exercise* 2.1.4*: *Write* $ay'' + by' + cy = f(x)$ *as a first order system of ODEs.*

Exercise* 2.1.5*: *Write* $x'' + y^2y' - x^3 = \sin(t)$*,* $y'' + (x' + y')^2 - x = 0$ *as a first order system of ODEs.*

Exercise* 2.1.6*: *Suppose two masses on carts on frictionless surface are at displacements* x_1 *and* x_2 *as in Example 2.1.3 on page 74. Suppose that a rocket applies force F in the positive direction on cart* x_1*. Set up the system of equations.*

Exercise* 2.1.7*: *Suppose the tanks are as in Example 2.1.2 on page 73, starting both at volume* V*, but now the rate of flow from tank 1 to tank 2 is* r_1*, and rate of flow from tank 2 to tank one is* r_2*. Notice that the volumes are now not constant. Set up the system of equations.*

Exercise* 2.1.101*: *Find the general solution to* $y_1' = 3y_1$*,* $y_2' = y_1 + y_2$*,* $y_3' = y_1 + y_3$*.*

Exercise* 2.1.102*: *Solve* $y' = 2x$*,* $x' = x + y$*,* $x(0) = 1$*,* $y(0) = 3$*.*

Exercise* 2.1.103*: *Write* $x''' = x + t$ *as a first order system.*

Exercise* 2.1.104*: *Write* $y_1'' + y_1 + y_2 = t$*,* $y_2'' + y_1 - y_2 = t^2$ *as a first order system.*

Exercise* 2.1.105*: *Suppose two masses on carts on frictionless surface are at displacements* x_1 *and* x_2 *as in Example 2.1.3 on page 74. Suppose initial displacement is* $x_1(0) = x_2(0) = 0$*, and initial velocity is* $x_1'(0) = x_2'(0) = a$ *for some number* a*. Use your intuition to solve the system, explain your reasoning.*

Exercise* 2.1.106*: *Suppose the tanks are as in Example 2.1.2 on page 73 except that clean water flows in at the rate* s *liters per second into tank 1, and brine flows out of tank 2 and into the sewer also at the rate of* s *liters per second. The rate of flow from tank 1 into tank 2 is still* r*, but the rate of flow from tank 2 back into tank 1 is* $r - s$ *(assume* $r > s$*).*

a) *Draw the picture.*
b) *Set up the system of equations.*
c) *Intuitively, what happens as* t *goes to infinity, explain.*

2.2 Matrices and linear systems

Note: 1.5 lectures, first part of §5.1 in [EP], §7.2 and §7.3 in [BD], see also appendix A

2.2.1 Matrices and vectors

Before we start talking about linear systems of ODEs, we need to talk about matrices, so let us review these briefly. A *matrix* is an $m \times n$ array of numbers (m rows and n columns). For example, we denote a 3×5 matrix as follows

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix}.$$

The numbers a_{ij} are called *elements* or *entries*.

By a *vector* we usually mean a *column vector*, that is an $m \times 1$ matrix. If we mean a *row vector*, we will explicitly say so (a row vector is a $1 \times n$ matrix). We usually denote matrices by upper case letters and vectors by lower case letters with an arrow such as $\vec{x}$ or $\vec{b}$. By $\vec{0}$ we mean the vector of all zeros.

We define some operations on matrices. We want 1×1 matrices to really act like numbers, so our operations have to be compatible with this viewpoint.

First, we can multiply a matrix by a *scalar* (a number). We simply multiply each entry in the matrix by the scalar. For example,

$$2 \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}.$$

Matrix addition is also easy. We add matrices element by element. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 2 \\ 4 & 7 & 10 \end{bmatrix}.$$

If the sizes do not match, then addition is not defined.

If we denote by 0 the matrix with all zero entries, by c, d scalars, and by A, B, C matrices, we have the following familiar rules:

$$\begin{aligned} A + 0 = A &= 0 + A, \\ A + B &= B + A, \\ (A + B) + C &= A + (B + C), \\ c(A + B) &= cA + cB, \\ (c + d)A &= cA + dA. \end{aligned}$$

Another useful operation for matrices is the so-called *transpose*. This operation just swaps rows and columns of a matrix. The transpose of A is denoted by A^T. Example:

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

2.2.2 Matrix multiplication

Let us now define matrix multiplication. First we define the so-called *dot product* (or *inner product*) of two vectors. Usually this will be a row vector multiplied with a column vector of the same size. For the dot product we multiply each pair of entries from the first and the second vector and we sum these products. The result is a single number. For example,

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1b_1 + a_2b_2 + a_3b_3.$$

Similarly for larger (or smaller) vectors.

Armed with the dot product we define the *product of matrices*. First let us denote by $\operatorname{row}_i(A)$ the i^{th} row of A and by $\operatorname{column}_j(A)$ the j^{th} column of A. For an $m \times n$ matrix A and an $n \times p$ matrix B, we can define the product AB. We let AB be an $m \times p$ matrix whose ij^{th} entry is the dot product

$$\operatorname{row}_i(A) \cdot \operatorname{column}_j(B).$$

Do note how the sizes match up: $m \times n$ multiplied by $n \times p$ is $m \times p$. Example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 0 & 1 \cdot (-1) + 2 \cdot 1 + 3 \cdot 0 \\ 4 \cdot 1 + 5 \cdot 1 + 6 \cdot 1 & 4 \cdot 0 + 5 \cdot 1 + 6 \cdot 0 & 4 \cdot (-1) + 5 \cdot 1 + 6 \cdot 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 1 \\ 15 & 5 & 1 \end{bmatrix}$$

For multiplication we want an analogue of a 1. This analogue is the so-called *identity matrix*. The identity matrix is a square matrix with 1s on the diagonal and zeros everywhere else. It is usually denoted by I. For each size we have a different identity matrix and so sometimes we may denote the size as a subscript. For example, the I_3 would be the 3×3 identity matrix

$$I = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We have the following rules for matrix multiplication. Suppose that A, B, C are matrices of the correct sizes so that the following make sense. Let α denote a scalar (number).

$$\begin{aligned} A(BC) &= (AB)C, \\ A(B+C) &= AB + AC, \\ (B+C)A &= BA + CA, \\ \alpha(AB) &= (\alpha A)B = A(\alpha B), \\ IA &= A = AI. \end{aligned}$$

A few warnings are in order.

(i) $AB \neq BA$ in general (it may be true by fluke sometimes). That is, matrices do not commute. For example, take $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right]$.

(ii) $AB = AC$ does not necessarily imply $B = C$, even if A is not 0.

(iii) $AB = 0$ does not necessarily mean that $A = 0$ or $B = 0$. Try, for example, $A = B = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$.

For the last two items to hold we would need to "divide" by a matrix. This is where the *matrix inverse* comes in. Suppose that A and B are $n \times n$ matrices such that

$$AB = I = BA.$$

Then we call B the inverse of A and we denote B by A^{-1}. If the inverse of A exists, then we call A *invertible*. If A is not invertible, we sometimes say A is *singular*.

If A is invertible, then $AB = AC$ does imply that $B = C$ (in particular, the inverse of A is unique). We just multiply both sides by A^{-1} (on the left) to get $A^{-1}AB = A^{-1}AC$ or $IB = IC$ or $B = C$. It is also not hard to see that $(A^{-1})^{-1} = A$.

2.2.3 The determinant

For square matrices we define a useful quantity called the *determinant*. We define the determinant of a 1×1 matrix as the value of its only entry. For a 2×2 matrix we define

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \overset{\text{def}}{=} ad - bc.$$

Before trying to define the determinant for larger matrices, let us note the meaning of the determinant. Consider an $n \times n$ matrix as a mapping of the n-dimensional euclidean space $\mathbb{R}^n$ to itself, where $\vec{x}$ gets sent to $A\vec{x}$. In particular, a 2×2 matrix A is a mapping of the plane to itself. The determinant of A is the factor by which the area of objects changes. If we take the unit square (square of side 1) in the plane, then A takes the square to a parallelogram of area $|\det(A)|$. The sign of $\det(A)$ denotes changing of orientation (negative if the axes get flipped). For example, let

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

Then $\det(A) = 1 + 1 = 2$. Let us see where the (unit) square with vertices $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$ gets sent. Clearly $(0,0)$ gets sent to $(0,0)$.

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

The image of the square is another square with vertices $(0,0)$, $(1,-1)$, $(1,1)$, and $(2,0)$. The image square has a side of length $\sqrt{2}$ and is therefore of area 2.

If you think back to high school geometry, you may have seen a formula for computing the area of a parallelogram with vertices $(0,0)$, (a,c), (b,d) and $(a+b, c+d)$. And it is precisely

$$\left| \det \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) \right|.$$

The vertical lines above mean absolute value. The matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ carries the unit square to the given parallelogram.

Let us look at the determinant for larger matrices. We define A_{ij} as the matrix A with the i^{th} row and the j^{th} column deleted. To compute the determinant of a matrix, pick one row, say the i^{th} row and compute:

$$\boxed{\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).}$$

For the first row we get

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13}) - \cdots \begin{cases} +a_{1n} \det(A_{1n}) & \text{if } n \text{ is odd,} \\ -a_{1n} \det(A_{1n}) & \text{if } n \text{ even.} \end{cases}$$

We alternately add and subtract the determinants of the submatrices A_{ij} multiplied by a_{ij} for a fixed i and all j. For a 3×3 matrix, picking the first row, we get $\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13})$. For example,

$$\begin{aligned} \det \left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \right) &= 1 \cdot \det \left(\begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} \right) - 2 \cdot \det \left(\begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} \right) + 3 \cdot \det \left(\begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} \right) \\ &= 1(5 \cdot 9 - 6 \cdot 8) - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) = 0. \end{aligned}$$

The numbers $(-1)^{i+j} \det(A_{ij})$ are called *cofactors* of the matrix and this way of computing the determinant is called the *cofactor expansion*. No matter which row you pick, you always get the same number. It is also possible to compute the determinant by expanding along columns (picking a column instead of a row above). It is true that $\det(A) = \det(A^T)$.

A common notation for the determinant is a pair of vertical lines:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right).$$

I personally find this notation confusing as vertical lines usually mean a positive quantity, while determinants can be negative. Also think about how to write the absolute value of a determinant. I will not use this notation in this book.

Think of the determinants telling you the scaling of a mapping. If B doubles the sizes of geometric objects and A triples them, then AB (which applies B to an object and then A) should make size go up by a factor of 6. This is true in general:

$$\det(AB) = \det(A)\det(B).$$

This property is one of the most useful, and it is employed often to actually compute determinants. A particularly interesting consequence is to note what it means for existence of inverses. Take A and B to be inverses of each other, that is $AB = I$. Then

$$\det(A)\det(B) = \det(AB) = \det(I) = 1.$$

Neither $\det(A)$ nor $\det(B)$ can be zero. Let us state this as a theorem as it will be very important in the context of this course.

Theorem 2.2.1. *An $n \times n$ matrix A is invertible if and only if* $\det(A) \neq 0$.

In fact, $\det(A^{-1})\det(A) = 1$ says that $\det(A^{-1}) = \frac{1}{\det(A)}$. So we even know what the determinant of A^{-1} is before we know how to compute A^{-1}.

There is a simple formula for the inverse of a 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Notice the determinant of the matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ in the denominator of the fraction. The formula only works if the determinant is nonzero, otherwise we are dividing by zero.

2.2.4 Solving linear systems

One application of matrices we will need is to solve systems of linear equations. This is best shown by example. Suppose that we have the following system of linear equations

$$\begin{aligned} 2x_1 + 2x_2 + 2x_3 &= 2, \\ x_1 + x_2 + 3x_3 &= 5, \\ x_1 + 4x_2 + x_3 &= 10. \end{aligned}$$

Without changing the solution, we could swap equations in this system, we could multiply any of the equations by a nonzero number, and we could add a multiple of one equation to another equation. It turns out these operations always suffice to find a solution.

It is easier to write the system as a matrix equation. The system above can be written as

$$\begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix}.$$

To solve the system we put the coefficient matrix (the matrix on the left-hand side of the equation) together with the vector on the right and side and get the so-called *augmented matrix*

$$\left[\begin{array}{ccc|c} 2 & 2 & 2 & 2 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{array}\right].$$

We apply the following three elementary operations.

(i) Swap two rows.

(ii) Multiply a row by a nonzero number.

(iii) Add a multiple of one row to another row.

We keep doing these operations until we get into a state where it is easy to read off the answer, or until we get into a contradiction indicating no solution, for example if we come up with an equation such as $0 = 1$.

Let us work through the example. First multiply the first row by $1/2$ to obtain

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{array}\right].$$

Now subtract the first row from the second and third row.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & 3 & 0 & 9 \end{array}\right]$$

Multiply the last row by $1/3$ and the second row by $1/2$.

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

Swap rows 2 and 3.

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

Subtract the last row from the first, then subtract the second row from the first.

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

If we think about what equations this augmented matrix represents, we see that $x_1 = -4$, $x_2 = 3$, and $x_3 = 2$. We try this solution in the original system and, voilà, it works!

***Exercise* 2.2.1:** *Check that the solution above really solves the given equations.*

We write this equation in matrix notation as

$$A\vec{x} = \vec{b},$$

where A is the matrix $\begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix}$ and $\vec{b}$ is the vector $\begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix}$. The solution can also be computed via the inverse,

$$\vec{x} = A^{-1}A\vec{x} = A^{-1}\vec{b}.$$

It is possible that the solution is not unique, or that no solution exists. It is easy to tell if a solution does not exist. If during the row reduction you come up with a row where all the entries except the last one are zero (the last entry in a row corresponds to the right-hand side of the equation), then the system is *inconsistent* and has no solution. For example, for a system of 3 equations and 3 unknowns, if you find a row such as $[\,0 \quad 0 \quad 0 \mid 1\,]$ in the augmented matrix, you know the system is inconsistent. That row corresponds to $0 = 1$.

You generally try to use row operations until the following conditions are satisfied. The first (from the left) nonzero entry in each row is called the *leading entry*.

(i) The leading entry in any row is strictly to the right of the leading entry of the row above.

(ii) Any zero rows are below all the nonzero rows.

(iii) All leading entries are 1.

(iv) All the entries above and below a leading entry are zero.

Such a matrix is said to be in *reduced row echelon form*. The variables corresponding to columns with no leading entries are said to be *free variables*. Free variables mean that we can pick those variables to be anything we want and then solve for the rest of the unknowns.

Example 2.2.1: The following augmented matrix is in reduced row echelon form.

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

Suppose the variables are x_1, x_2, and x_3. Then x_2 is the free variable, $x_1 = 3 - 2x_2$, and $x_3 = 1$.

On the other hand if during the row reduction process you come up with the matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 13 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{array}\right],$$

there is no need to go further. The last row corresponds to the equation $0x_1 + 0x_2 + 0x_3 = 3$, which is preposterous. Hence, no solution exists.

2.2.5 Computing the inverse

If the matrix A is square and there exists a unique solution $\vec{x}$ to $A\vec{x} = \vec{b}$ for any $\vec{b}$ (there are no free variables), then A is invertible. Multiplying both sides by A^{-1}, you can see that $\vec{x} = A^{-1}\vec{b}$. So it is useful to compute the inverse if you want to solve the equation for many different right-hand sides $\vec{b}$.

We have a formula for the 2×2 inverse, but it is also not hard to compute inverses of larger matrices. While we will not have too much occasion to compute inverses for larger matrices than 2×2 by hand, let us touch on how to do it. Finding the inverse of A is actually just solving a bunch of linear equations. If we can solve $A\vec{x}_k = \vec{e}_k$ where $\vec{e}_k$ is the vector with all zeros except a 1 at the k^{th} position, then the inverse is the matrix with the columns $\vec{x}_k$ for $k = 1, 2, \ldots, n$ (exercise: why?). Therefore, to find the inverse we write a larger $n \times 2n$ augmented matrix $[\, A \mid I \,]$, where I is the identity matrix. We then perform row reduction. The reduced row echelon form of $[\, A \mid I \,]$ will be of the form $[\, I \mid A^{-1} \,]$ if and only if A is invertible. We then just read off the inverse A^{-1}.

2.2.6 Exercises

Exercise* 2.2.2*: *Solve* $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \vec{x} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$ *by using matrix inverse.*

Exercise* 2.2.3*: *Compute determinant of* $\begin{bmatrix} 9 & -2 & -6 \\ -8 & 3 & 6 \\ 10 & -2 & -6 \end{bmatrix}$.

Exercise* 2.2.4*: *Compute determinant of* $\begin{bmatrix} 1 & 2 & 3 & 1 \\ 4 & 0 & 5 & 0 \\ 6 & 0 & 7 & 0 \\ 8 & 0 & 10 & 1 \end{bmatrix}$. *Hint: Expand along the proper row or column to make the calculations simpler.*

Exercise* 2.2.5*: *Compute inverse of* $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Exercise* 2.2.6*: *For which h is* $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & h \end{bmatrix}$ *not invertible? Is there only one such h? Are there several? Infinitely many?*

Exercise* 2.2.7*: *For which h is* $\begin{bmatrix} h & 1 & 1 \\ 0 & h & 0 \\ 1 & 1 & h \end{bmatrix}$ *not invertible? Find all such h.*

Exercise* 2.2.8*: *Solve* $\begin{bmatrix} 9 & -2 & -6 \\ -8 & 3 & 6 \\ 10 & -2 & -6 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Exercise* 2.2.9*: *Solve* $\begin{bmatrix} 5 & 3 & 7 \\ 8 & 4 & 4 \\ 6 & 3 & 3 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$.

Exercise* 2.2.10*: *Solve* $\begin{bmatrix} 3 & 2 & 3 & 0 \\ 3 & 3 & 3 & 3 \\ 0 & 2 & 4 & 2 \\ 2 & 3 & 4 & 3 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 0 \\ 4 \\ 1 \end{bmatrix}$.

Exercise* 2.2.11*: *Find 3 nonzero 2×2 matrices A, B, and C such that $AB = AC$ but $B \neq C$.*

Exercise* 2.2.101*: *Compute determinant of* $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & -5 \\ 1 & -1 & 0 \end{bmatrix}$

Exercise* 2.2.102*: *Find t such that $\left[\begin{smallmatrix} 1 & t \\ -1 & 2 \end{smallmatrix}\right]$ is not invertible.*

Exercise* 2.2.103*: *Solve $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \vec{x} = \left[\begin{smallmatrix} 10 \\ 20 \end{smallmatrix}\right]$.*

Exercise* 2.2.104*: *Suppose a, b, c are nonzero numbers. Let $M = \left[\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right]$, $N = \left[\begin{smallmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{smallmatrix}\right]$.*

a) Compute M^{-1}.

b) Compute N^{-1}.

2.3 Linear systems of ODEs

Note: less than 1 lecture, second part of §5.1 in [EP], §7.4 in [BD]

First let us talk about matrix- or vector-valued functions. Such a function is just a matrix or vector whose entries depend on some variable. If t is the independent variable, we write a *vector-valued function* $\vec{x}(t)$ as

$$\vec{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

Similarly a *matrix-valued function* $A(t)$ is

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{bmatrix}.$$

The derivative $A'(t)$ or $\frac{dA}{dt}$ is just the matrix-valued function whose ij^{th} entry is $a'_{ij}(t)$.

Rules of differentiation of matrix-valued functions are similar to rules for normal functions. Let $A(t)$ and $B(t)$ be matrix-valued functions. Let c a scalar and let C be a constant matrix. Then

$$\begin{aligned} \big(A(t) + B(t)\big)' &= A'(t) + B'(t), \\ \big(A(t)B(t)\big)' &= A'(t)B(t) + A(t)B'(t), \\ \big(cA(t)\big)' &= cA'(t), \\ \big(CA(t)\big)' &= CA'(t), \\ \big(A(t)\,C\big)' &= A'(t)\,C. \end{aligned}$$

Note the order of the multiplication in the last two expressions.

A *first order linear system of ODEs* is a system that can be written as the vector equation

$$\vec{x}'(t) = P(t)\vec{x}(t) + \vec{f}(t),$$

where $P(t)$ is a matrix-valued function, and $\vec{x}(t)$ and $\vec{f}(t)$ are vector-valued functions. We will often suppress the dependence on t and only write $\vec{x}' = P\vec{x} + \vec{f}$. A solution of the system is a vector-valued function $\vec{x}$ satisfying the vector equation.

For example, the equations

$$\begin{aligned} x_1' &= 2t x_1 + e^t x_2 + t^2, \\ x_2' &= \frac{x_1}{t} - x_2 + e^t, \end{aligned}$$

can be written as

$$\vec{x}' = \begin{bmatrix} 2t & e^t \\ 1/t & -1 \end{bmatrix} \vec{x} + \begin{bmatrix} t^2 \\ e^t \end{bmatrix}.$$

We will mostly concentrate on equations that are not just linear, but are in fact *constant coefficient* equations. That is, the matrix P will be constant; it will not depend on t.

When $\vec{f} = \vec{0}$ (the zero vector), then we say the system is *homogeneous*. For homogeneous linear systems we have the principle of superposition, just like for single homogeneous equations.

Theorem 2.3.1 (Superposition)**.** *Let $\vec{x}' = P\vec{x}$ be a linear homogeneous system of ODEs. Suppose that $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are n solutions of the equation and $c_1, c_2, \ldots, c_n$ are any constants, then*

$$\vec{x} = c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_n\vec{x}_n, \tag{2.2}$$

is also a solution. Furthermore, if this is a system of n equations (P is $n \times n$), and $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are linearly independent, then every solution $\vec{x}$ can be written as (2.2)*.*

Linear independence for vector-valued functions is the same idea as for normal functions. The vector-valued functions $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are linearly independent when

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_n\vec{x}_n = \vec{0}$$

has only the solution $c_1 = c_2 = \cdots = c_n = 0$, where the equation must hold for all t.

Example 2.3.1: $\vec{x}_1 = \begin{bmatrix} t^2 \\ t \end{bmatrix}$, $\vec{x}_2 = \begin{bmatrix} 0 \\ 1+t \end{bmatrix}$, $\vec{x}_3 = \begin{bmatrix} -t^2 \\ 1 \end{bmatrix}$ are linearly dependent because $\vec{x}_1 + \vec{x}_3 = \vec{x}_2$, and this holds for all t. So $c_1 = 1$, $c_2 = -1$, and $c_3 = 1$ above will work.

On the other hand if we change the example just slightly $\vec{x}_1 = \begin{bmatrix} t^2 \\ t \end{bmatrix}$, $\vec{x}_2 = \begin{bmatrix} 0 \\ t \end{bmatrix}$, $\vec{x}_3 = \begin{bmatrix} -t^2 \\ 1 \end{bmatrix}$, then the functions are linearly independent. First write $c_1\vec{x}_1 + c_2\vec{x}_2 + c_3\vec{x}_3 = \vec{0}$ and note that it has to hold for all t. We get that

$$c_1\vec{x}_1 + c_2\vec{x}_2 + c_3\vec{x}_3 = \begin{bmatrix} c_1t^2 - c_3t^2 \\ c_1t + c_2t + c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In other words $c_1t^2 - c_3t^2 = 0$ and $c_1t + c_2t + c_3 = 0$. If we set $t = 0$, then the second equation becomes $c_3 = 0$. But then the first equation becomes $c_1t^2 = 0$ for all t and so $c_1 = 0$. Thus the second equation is just $c_2t = 0$, which means $c_2 = 0$. So $c_1 = c_2 = c_3 = 0$ is the only solution and $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ are linearly independent.

The linear combination $c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_n\vec{x}_n$ could always be written as

$$X(t)\,\vec{c},$$

where $X(t)$ is the matrix with columns $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$, and $\vec{c}$ is the column vector with entries $c_1, c_2, \ldots, c_n$. Assuming that $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are linearly independent, the matrix-valued function $X(t)$ is called a *fundamental matrix*, or a *fundamental matrix solution*.

To solve nonhomogeneous first order linear systems, we use the same technique as we applied to solve single linear nonhomogeneous equations.

Theorem 2.3.2. *Let $\vec{x}' = P\vec{x} + \vec{f}$ be a linear system of ODEs. Suppose $\vec{x}_p$ is one particular solution. Then every solution can be written as*

$$\vec{x} = \vec{x}_c + \vec{x}_p,$$

where $\vec{x}_c$ is a solution to the associated homogeneous equation ($\vec{x}' = P\vec{x}$).

The procedure for systems is the same as for single equations. We find a particular solution to the nonhomogeneous equation, then we find the general solution to the associated homogeneous equation, and finally we add the two together.

Alright, suppose you have found the general solution of $\vec{x}' = P\vec{x} + \vec{f}$. Next suppose you are given an initial condition of the form

$$\vec{x}(t_0) = \vec{b}$$

for some fixed t_0 and a constant vector $\vec{b}$. Let $X(t)$ be a fundamental matrix solution of the associated homogeneous equation (i.e. columns of $X(t)$ are solutions). The general solution can be written as

$$\vec{x}(t) = X(t)\,\vec{c} + \vec{x}_p(t).$$

We are seeking a vector $\vec{c}$ such that

$$\vec{b} = \vec{x}(t_0) = X(t_0)\,\vec{c} + \vec{x}_p(t_0).$$

In other words, we are solving for $\vec{c}$ the nonhomogeneous system of linear equations

$$X(t_0)\,\vec{c} = \vec{b} - \vec{x}_p(t_0).$$

Example 2.3.2: In § 2.1 we solved the system

$$\begin{aligned} x_1' &= x_1, \\ x_2' &= x_1 - x_2, \end{aligned}$$

with initial conditions $x_1(0) = 1$, $x_2(0) = 2$. Let us consider this problem in the language of this section.

The system is homogeneous, so $\vec{f}(t) = \vec{0}$. We write the system and the initial conditions as

$$\vec{x}' = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \vec{x}, \qquad \vec{x}(0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We found the general solution is $x_1 = c_1 e^t$ and $x_2 = \frac{c_1}{2} e^t + c_2 e^{-t}$. Letting $c_1 = 1$ and $c_2 = 0$, we obtain the solution $\left[\begin{smallmatrix} e^t \\ (1/2)e^t \end{smallmatrix} \right]$. Letting $c_1 = 0$ and $c_2 = 1$, we obtain $\left[\begin{smallmatrix} 0 \\ e^{-t} \end{smallmatrix} \right]$. These two solutions are linearly independent, as can be seen by setting $t = 0$, and noting that the resulting constant vectors are linearly independent. In matrix notation, a fundamental matrix solution is, therefore,

$$X(t) = \begin{bmatrix} e^t & 0 \\ \frac{1}{2} e^t & e^{-t} \end{bmatrix}.$$

To solve the initial value problem we solve for $\vec{c}$ in the equation

$$X(0)\,\vec{c} = \vec{b},$$

or in other words,

$$\begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \vec{c} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

A single elementary row operation shows $\vec{c} = \left[\begin{smallmatrix} 1 \\ 3/2 \end{smallmatrix}\right]$. Our solution is

$$\vec{x}(t) = X(t)\,\vec{c} = \begin{bmatrix} e^t & 0 \\ \frac{1}{2}e^t & e^{-t} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{3}{2} \end{bmatrix} = \begin{bmatrix} e^t \\ \frac{1}{2}e^t + \frac{3}{2}e^{-t} \end{bmatrix}.$$

This new solution agrees with our previous solution from § 2.1.

2.3.1 Exercises

Exercise* 2.3.1*: *Write the system* $x_1' = 2x_1 - 3tx_2 + \sin t$, $x_2' = e^t x_1 + 3x_2 + \cos t$ *in the form* $\vec{x}' = P(t)\vec{x} + \vec{f}(t)$.

Exercise* 2.3.2*:

a) *Verify that the system* $\vec{x}' = \left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right] \vec{x}$ *has the two solutions* $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] e^{4t}$ *and* $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] e^{-2t}$.

b) *Write down the general solution.*

c) *Write down the general solution in the form* $x_1 =?$, $x_2 =?$ *(i.e. write down a formula for each element of the solution).*

Exercise* 2.3.3*: *Verify that* $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] e^t$ *and* $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] e^t$ *are linearly independent. Hint: Just plug in* $t = 0$.

Exercise* 2.3.4*: *Verify that* $\left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right] e^t$ *and* $\left[\begin{smallmatrix} 1 \\ -1 \\ 1 \end{smallmatrix}\right] e^t$ *and* $\left[\begin{smallmatrix} 1 \\ -1 \\ 1 \end{smallmatrix}\right] e^{2t}$ *are linearly independent. Hint: You must be a bit more tricky than in the previous exercise.*

Exercise* 2.3.5*: *Verify that* $\left[\begin{smallmatrix} t \\ t^2 \end{smallmatrix}\right]$ *and* $\left[\begin{smallmatrix} t^3 \\ t^4 \end{smallmatrix}\right]$ *are linearly independent.*

Exercise* 2.3.6*: *Take the system* $x_1' + x_2' = x_1$, $x_1' - x_2' = x_2$.

a) *Write it in the form* $A\vec{x}' = B\vec{x}$ *for matrices* A *and* B.

b) *Compute* A^{-1} *and use that to write the system in the form* $\vec{x}' = P\vec{x}$.

Exercise* 2.3.101*: *Are* $\left[\begin{smallmatrix} e^{2t} \\ e^t \end{smallmatrix}\right]$ *and* $\left[\begin{smallmatrix} e^t \\ e^{2t} \end{smallmatrix}\right]$ *linearly independent? Justify.*

Exercise* 2.3.102*: *Are* $\left[\begin{smallmatrix} \cosh(t) \\ 1 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} e^t \\ 1 \end{smallmatrix}\right]$, *and* $\left[\begin{smallmatrix} e^{-t} \\ 1 \end{smallmatrix}\right]$ *linearly independent? Justify.*

Exercise* 2.3.103*: *Write* $x' = 3x - y + e^t$, $y' = tx$ *in matrix notation.*

Exercise* 2.3.104*:

a) *Write* $x_1' = 2tx_2$, $x_2' = 2tx_2$ *in matrix notation.*

b) *Solve and write the solution in matrix notation.*

2.4 Eigenvalue method

Note: 2 lectures, §5.2 in [EP], part of §7.3, §7.5, and §7.6 in [BD]

In this section we will learn how to solve linear homogeneous constant coefficient systems of ODEs by the eigenvalue method. Suppose we have such a system

$$\vec{x}' = P\vec{x},$$

where P is a constant square matrix. We wish to adapt the method for the single constant coefficient equation by trying the function $e^{\lambda t}$. However, $\vec{x}$ is a vector. So we try $\vec{x} = \vec{v}e^{\lambda t}$, where $\vec{v}$ is an arbitrary constant vector. We plug this $\vec{x}$ into the equation to get

$$\underbrace{\lambda\vec{v}e^{\lambda t}}_{\vec{x}'} = \underbrace{P\vec{v}e^{\lambda t}}_{P\vec{x}}.$$

We divide by $e^{\lambda t}$ and notice that we are looking for a scalar λ and a vector $\vec{v}$ that satisfy the equation

$$\lambda\vec{v} = P\vec{v}.$$

To solve this equation we need a little bit more linear algebra, which we now review.

2.4.1 Eigenvalues and eigenvectors of a matrix

Let A be a constant square matrix. Suppose there is a scalar λ and a nonzero vector $\vec{v}$ such that

$$A\vec{v} = \lambda\vec{v}.$$

We call λ an *eigenvalue* of A and we call $\vec{v}$ a corresponding *eigenvector*.

Example 2.4.1: The matrix $\left[\begin{smallmatrix} 2 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ has an eigenvalue $\lambda = 2$ with a corresponding eigenvector $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ as

$$\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Let us see how to compute eigenvalues for any matrix. Rewrite the equation for an eigenvalue as

$$(A - \lambda I)\vec{v} = \vec{0}.$$

This equation has a nonzero solution $\vec{v}$ only if $A - \lambda I$ is not invertible. Were it invertible, we could write $(A - \lambda I)^{-1}(A - \lambda I)\vec{v} = (A - \lambda I)^{-1}\vec{0}$, which implies $\vec{v} = \vec{0}$. Therefore, A has the eigenvalue λ if and only if λ solves the equation

$$\det(A - \lambda I) = 0.$$

Consequently, we can find an eigenvalue of A without finding a corresponding eigenvector at the same time. An eigenvector will have to be found later, once λ is known.

Example 2.4.2: Find all eigenvalues of $\left[\begin{smallmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right]$.

We write

$$\det\left(\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{bmatrix}\right) =$$
$$= (2-\lambda)\big((2-\lambda)^2 - 1\big) = -(\lambda-1)(\lambda-2)(\lambda-3).$$

So the eigenvalues are $\lambda = 1$, $\lambda = 2$, and $\lambda = 3$.

For an $n \times n$ matrix, the polynomial we get by computing $\det(A - \lambda I)$ is of degree n, and hence in general, we have n eigenvalues. Some may be repeated, some may be complex.

To find an eigenvector corresponding to an eigenvalue λ, we write

$$(A - \lambda I)\vec{v} = \vec{0},$$

and solve for a nontrivial (nonzero) vector $\vec{v}$. If λ is an eigenvalue, there will be at least one free variable, and so for each distinct eigenvalue λ, we can always find an eigenvector.

Example 2.4.3: Find an eigenvector of $\left[\begin{smallmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right]$ corresponding to the eigenvalue $\lambda = 3$.

We write

$$(A - \lambda I)\vec{v} = \left(\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} - 3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \vec{0}.$$

It is easy to solve this system of linear equations. We write down the augmented matrix

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

and perform row operations (exercise: which ones?) until we get:

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

The entries of $\vec{v}$ have to satisfy the equations $v_1 - v_2 = 0$, $v_3 = 0$, and v_2 is a free variable. We can pick v_2 to be arbitrary (but nonzero), let $v_1 = v_2$, and of course $v_3 = 0$. For example, if we pick $v_2 = 1$, then $\vec{v} = \left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right]$. Let us verify that $\vec{v}$ really is an eigenvector corresponding to $\lambda = 3$:

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Yay! It worked.

Exercise* 2.4.1** (easy):*** *Are eigenvectors unique? Can you find a different eigenvector for $\lambda = 3$ in the example above? How are the two eigenvectors related?*

Exercise* 2.4.2*: *When the matrix is 2×2 you do not need to do row operations when computing an eigenvector, you can read it off from $A - \lambda I$ (if you have computed the eigenvalues correctly). Can you see why? Explain. Try it for the matrix $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$.*

2.4.2 The eigenvalue method with distinct real eigenvalues

OK. We have the system of equations

$$\vec{x}' = P\vec{x}.$$

We find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the matrix P, and corresponding eigenvectors $\vec{v}_1$, $\vec{v}_2, \ldots, \vec{v}_n$. Now we notice that the functions $\vec{v}_1 e^{\lambda_1 t}, \vec{v}_2 e^{\lambda_2 t}, \ldots, \vec{v}_n e^{\lambda_n t}$ are solutions of the system of equations and hence $\vec{x} = c_1 \vec{v}_1 e^{\lambda_1 t} + c_2 \vec{v}_2 e^{\lambda_2 t} + \cdots + c_n \vec{v}_n e^{\lambda_n t}$ is a solution.

Theorem 2.4.1. *Take $\vec{x}' = P\vec{x}$. If P is an $n \times n$ constant matrix that has n distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then there exist n linearly independent corresponding eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and the general solution to $\vec{x}' = P\vec{x}$ can be written as*

$$\boxed{\vec{x} = c_1 \vec{v}_1 e^{\lambda_1 t} + c_2 \vec{v}_2 e^{\lambda_2 t} + \cdots + c_n \vec{v}_n e^{\lambda_n t}.}$$

The corresponding fundamental matrix solution is

$$X(t) = \begin{bmatrix} \vec{v}_1 e^{\lambda_1 t} & \vec{v}_2 e^{\lambda_2 t} & \cdots & \vec{v}_n e^{\lambda_n t} \end{bmatrix}.$$

That is, $X(t)$ is the matrix whose j^{th} column is $\vec{v}_j e^{\lambda_j t}$.

Example 2.4.4: Consider the system

$$\vec{x}' = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \vec{x}.$$

Find the general solution.

Earlier, we found the eigenvalues are $1, 2, 3$. We found the eigenvector $\left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right]$ for the eigenvalue 3. Similarly we find the eigenvector $\left[\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right]$ for the eigenvalue 1, and $\left[\begin{smallmatrix} 0 \\ 1 \\ -1 \end{smallmatrix}\right]$ for the eigenvalue 2 (exercise: check). Hence our general solution is

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} e^{t} + c_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} e^{2t} + c_3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} e^{3t} = \begin{bmatrix} c_1 e^{t} + c_3 e^{3t} \\ -c_1 e^{t} + c_2 e^{2t} + c_3 e^{3t} \\ -c_2 e^{2t} \end{bmatrix}.$$

In terms of a fundamental matrix solution,

$$\vec{x} = X(t)\,\vec{c} = \begin{bmatrix} e^{t} & 0 & e^{3t} \\ -e^{t} & e^{2t} & e^{3t} \\ 0 & -e^{2t} & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

***Exercise* 2.4.3:** *Check that this $\vec{x}$ really solves the system.*

Note: If we write a single homogeneous linear constant coefficient n^{th} order equation as a first order system (as we did in § 2.1), then the eigenvalue equation

$$\det(P - \lambda I) = 0$$

is essentially the same as the characteristic equation we got in § **??** and § **??**.

2.4.3 Complex eigenvalues

A matrix may very well have complex eigenvalues even if all the entries are real. Take, for example,

$$\vec{x}' = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \vec{x}.$$

Let us compute the eigenvalues of the matrix $P = \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$.

$$\det(P - \lambda I) = \det\left(\begin{bmatrix} 1-\lambda & 1 \\ -1 & 1-\lambda \end{bmatrix}\right) = (1-\lambda)^2 + 1 = \lambda^2 - 2\lambda + 2 = 0.$$

Thus $\lambda = 1 \pm i$. Corresponding eigenvectors are also complex. Start with $\lambda = 1 - i$.

$$\begin{aligned} \big(P - (1-i)I\big)\vec{v} &= \vec{0}, \\ \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \vec{v} &= \vec{0}. \end{aligned}$$

The equations $iv_1 + v_2 = 0$ and $-v_1 + iv_2 = 0$ are multiples of each other. So we only need to consider one of them. After picking $v_2 = 1$, for example, we have an eigenvector $\vec{v} = \left[\begin{smallmatrix} i \\ 1 \end{smallmatrix}\right]$. In similar fashion we find that $\left[\begin{smallmatrix} -i \\ 1 \end{smallmatrix}\right]$ is an eigenvector corresponding to the eigenvalue $1 + i$.

We could write the solution as

$$\vec{x} = c_1 \begin{bmatrix} i \\ 1 \end{bmatrix} e^{(1-i)t} + c_2 \begin{bmatrix} -i \\ 1 \end{bmatrix} e^{(1+i)t} = \begin{bmatrix} c_1 i e^{(1-i)t} - c_2 i e^{(1+i)t} \\ c_1 e^{(1-i)t} + c_2 e^{(1+i)t} \end{bmatrix}.$$

We would then need to look for complex values c_1 and c_2 to solve any initial conditions. It is perhaps not completely clear that we get a real solution. After solving for c_1 and c_2, we could use Euler's formula and do the whole song and dance we did before, but we will not. We will apply the formula in a smarter way first to find independent real solutions.

We claim that we did not have to look for a second eigenvector (nor for the second eigenvalue). All complex eigenvalues come in pairs (because the matrix P is real).

First a small detour. The real part of a complex number z can be computed as $\frac{z+\bar{z}}{2}$, where the bar above z means $\overline{a + ib} = a - ib$. This operation is called the *complex conjugate*. If a is a real number, then $\bar{a} = a$. Similarly we bar whole vectors or matrices by taking the complex conjugate of every entry. Suppose a matrix P is real. Then $\overline{P} = P$, and so $\overline{P\vec{x}} = \overline{P}\,\overline{\vec{x}} = P\overline{\vec{x}}$. Also the complex conjugate of 0 is still 0, therefore,

$$\vec{0} = \overline{\vec{0}} = \overline{(P - \lambda I)\vec{v}} = (P - \bar{\lambda} I)\overline{\vec{v}}.$$

In other words, if $\lambda = a + ib$ is an eigenvalue, then so is $\bar{\lambda} = a - ib$. And if $\vec{v}$ is an eigenvector corresponding to the eigenvalue λ, then $\overline{\vec{v}}$ is an eigenvector corresponding to the eigenvalue $\bar{\lambda}$.

Suppose $a + ib$ is a complex eigenvalue of P, and $\vec{v}$ is a corresponding eigenvector. Then

$$\vec{x}_1 = \vec{v} e^{(a+ib)t}$$

is a solution (complex-valued) of $\vec{x}' = P\vec{x}$. Euler's formula shows that $\overline{e^{a+ib}} = e^{a-ib}$, and so

$$\vec{x}_2 = \overline{\vec{x}_1} = \overline{\vec{v}} e^{(a-ib)t}$$

is also a solution. As $\vec{x}_1$ and $\vec{x}_2$ are solutions, the function

$$\vec{x}_3 = \operatorname{Re} \vec{x}_1 = \operatorname{Re} \vec{v} e^{(a+ib)t} = \frac{\vec{x}_1 + \overline{\vec{x}_1}}{2} = \frac{\vec{x}_1 + \vec{x}_2}{2} = \frac{1}{2}\vec{x}_1 + \frac{1}{2}\vec{x}_2$$

is also a solution. And $\vec{x}_3$ is real-valued! Similarly as $\operatorname{Im} z = \frac{z - \bar{z}}{2i}$ is the imaginary part, we find that

$$\vec{x}_4 = \operatorname{Im} \vec{x}_1 = \frac{\vec{x}_1 - \overline{\vec{x}_1}}{2i} = \frac{\vec{x}_1 - \vec{x}_2}{2i}.$$

is also a real-valued solution. It turns out that $\vec{x}_3$ and $\vec{x}_4$ are linearly independent. We will use Euler's formula to separate out the real and imaginary part.

Returning to our problem,

$$\vec{x}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix} e^{(1-i)t} = \begin{bmatrix} i \\ 1 \end{bmatrix} \left(e^t \cos t - i e^t \sin t \right) = \begin{bmatrix} i e^t \cos t + e^t \sin t \\ e^t \cos t - i e^t \sin t \end{bmatrix} = \begin{bmatrix} e^t \sin t \\ e^t \cos t \end{bmatrix} + i \begin{bmatrix} e^t \cos t \\ -e^t \sin t \end{bmatrix}.$$

Then

$$\operatorname{Re} \vec{x}_1 = \begin{bmatrix} e^t \sin t \\ e^t \cos t \end{bmatrix}, \qquad \text{and} \qquad \operatorname{Im} \vec{x}_1 = \begin{bmatrix} e^t \cos t \\ -e^t \sin t \end{bmatrix},$$

are the two real-valued linearly independent solutions we seek.

***Exercise* 2.4.4:** *Check that these really are solutions.*

The general solution is

$$\vec{x} = c_1 \begin{bmatrix} e^t \sin t \\ e^t \cos t \end{bmatrix} + c_2 \begin{bmatrix} e^t \cos t \\ -e^t \sin t \end{bmatrix} = \begin{bmatrix} c_1 e^t \sin t + c_2 e^t \cos t \\ c_1 e^t \cos t - c_2 e^t \sin t \end{bmatrix}.$$

This solution is real-valued for real c_1 and c_2. At this point, we would solve for any initial conditions we may have to find c_1 and c_2.

Let us summarize the discussion as a theorem.

Theorem 2.4.2. *Let P be a real-valued constant matrix. If P has a complex eigenvalue $a + ib$ and a corresponding eigenvector $\vec{v}$, then P also has a complex eigenvalue $a - ib$ with a corresponding eigenvector $\overline{\vec{v}}$. Furthermore, $\vec{x}' = P\vec{x}$ has two linearly independent real-valued solutions*

$$\vec{x}_1 = \operatorname{Re} \vec{v} e^{(a+ib)t}, \qquad \text{and} \qquad \vec{x}_2 = \operatorname{Im} \vec{v} e^{(a+ib)t}.$$

For each pair of complex eigenvalues $a + ib$ and $a - ib$, we get two real-valued linearly independent solutions. We then go on to the next eigenvalue, which is either a real eigenvalue or another complex eigenvalue pair. If we have n distinct eigenvalues (real or complex), then we end up with n linearly independent solutions. If we had only two equations ($n = 2$) as in the example above, then once we found two solutions we are finished, and our general solution is

$$\vec{x} = c_1\vec{x}_1 + c_2\vec{x}_2 = c_1\big(\operatorname{Re} \vec{v} e^{(a+ib)t}\big) + c_2\big(\operatorname{Im} \vec{v} e^{(a+ib)t}\big).$$

We can now find a real-valued general solution to any homogeneous system where the matrix has distinct eigenvalues. When we have repeated eigenvalues, matters get a bit more complicated and we will look at that situation in § 2.8.

2.4.4 Exercises

Exercise* 2.4.5** (easy):*** *Let A be a 3×3 matrix with an eigenvalue of 3 and a corresponding eigenvector $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$. Find $A\vec{v}$.*

Exercise* 2.4.6*:

a) *Find the general solution of $x_1' = 2x_1$, $x_2' = 3x_2$ using the eigenvalue method (first write the system in the form $\vec{x}' = A\vec{x}$).*

b) *Solve the system by solving each equation separately and verify you get the same general solution.*

Exercise* 2.4.7*: *Find the general solution of $x_1' = 3x_1 + x_2$, $x_2' = 2x_1 + 4x_2$ using the eigenvalue method.*

Exercise* 2.4.8*: *Find the general solution of $x_1' = x_1 - 2x_2$, $x_2' = 2x_1 + x_2$ using the eigenvalue method. Do not use complex exponentials in your solution.*

Exercise* 2.4.9*:

a) *Compute eigenvalues and eigenvectors of $A = \begin{bmatrix} 9 & -2 & -6 \\ -8 & 3 & 6 \\ 10 & -2 & -6 \end{bmatrix}$.*

b) *Find the general solution of $\vec{x}' = A\vec{x}$.*

Exercise* 2.4.10*: *Compute eigenvalues and eigenvectors of $\begin{bmatrix} -2 & -1 & -1 \\ 3 & 2 & 1 \\ -3 & -1 & 0 \end{bmatrix}$.*

Exercise* 2.4.11*: *Let a, b, c, d, e, f be numbers. Find the eigenvalues of* $\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$.

Exercise* 2.4.101*:

a) *Compute eigenvalues and eigenvectors of* $A = \begin{bmatrix} 1 & 0 & 3 \\ -1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix}$.

b) *Solve the system* $\vec{x}\,' = A\vec{x}$.

Exercise* 2.4.102*:

a) *Compute eigenvalues and eigenvectors of* $A = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$.

b) *Solve the system* $\vec{x}\,' = A\vec{x}$.

Exercise* 2.4.103*: *Solve* $x_1' = x_2$, $x_2' = x_1$ *using the eigenvalue method.*

Exercise* 2.4.104*: *Solve* $x_1' = x_2$, $x_2' = -x_1$ *using the eigenvalue method.*

2.5 Two-dimensional systems and their vector fields

Note: 1 lecture, part of §6.2 in [EP], parts of §7.5 and §7.6 in [BD]

Let us take a moment to talk about constant coefficient linear homogeneous systems in the plane. Much intuition can be obtained by studying this simple case. We use coordinates (x, y) for the plane as usual, and suppose $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a 2×2 matrix. Consider the system

$$\begin{bmatrix} x \\ y \end{bmatrix}' = P \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} x \\ y \end{bmatrix}' = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2.3}$$

The system is autonomous (compare this section to § 1.6) and so we can draw a vector field (see the end of § 2.1). We will be able to visually tell what the vector field looks like and how the solutions behave, once we find the eigenvalues and eigenvectors of the matrix P. For this section, we assume that P has two eigenvalues and two corresponding eigenvectors.

Case 1. Suppose that the eigenvalues of P are real and positive. We find two corresponding eigenvectors and plot them in the plane. For example, take the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. The eigenvalues are 1 and 2 and corresponding eigenvectors are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. See Figure 2.4.

Let (x, y) be a point on the line determined by an eigenvector $\vec{v}$ for an eigenvalue λ. That is, $\begin{bmatrix} x \\ y \end{bmatrix} = \alpha \vec{v}$ for some scalar α. Then

$$\begin{bmatrix} x \\ y \end{bmatrix}' = P \begin{bmatrix} x \\ y \end{bmatrix} = P(\alpha \vec{v}) = \alpha (P \vec{v}) = \alpha \lambda \vec{v}.$$

The derivative is a multiple of $\vec{v}$ and hence points along the line determined by $\vec{v}$. As $\lambda > 0$, the derivative points in the direction of $\vec{v}$ when α is positive and in the opposite direction when α is negative. We draw the lines determined by the eigenvectors, and we draw arrows on the lines to indicate the directions. See Figure 2.5 on the following page.

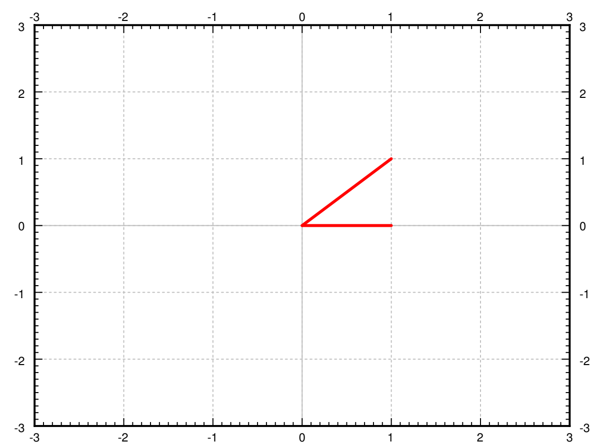


Figure 2.4: Eigenvectors of P.

We fill in the rest of the arrows for the vector field and we also draw a few solutions. See Figure 2.6 on the next page. The picture looks like a source with arrows coming out from the origin. Hence we call this type of picture a *source* or sometimes an *unstable node*.

Case 2. Suppose both eigenvalues are negative. For example, take the negation of the matrix in case 1, $\begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix}$. The eigenvalues are -1 and -2 and corresponding eigenvectors are the same, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. The calculation and the picture are almost the same. The only difference is that the eigenvalues are negative and hence all arrows are reversed. We get the picture in Figure 2.7 on the following page. We call this kind of picture a *sink* or a *stable node*.

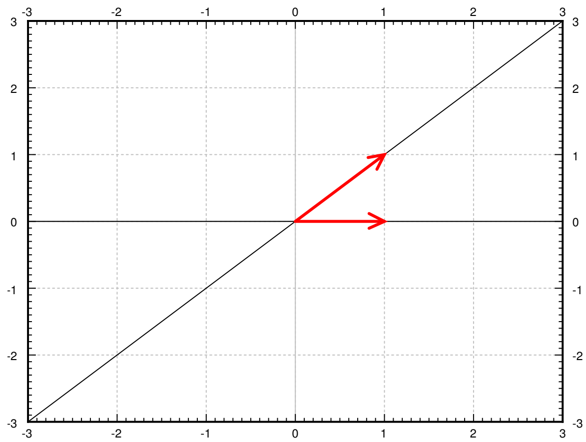


Figure 2.5: Eigenvectors of P with directions.

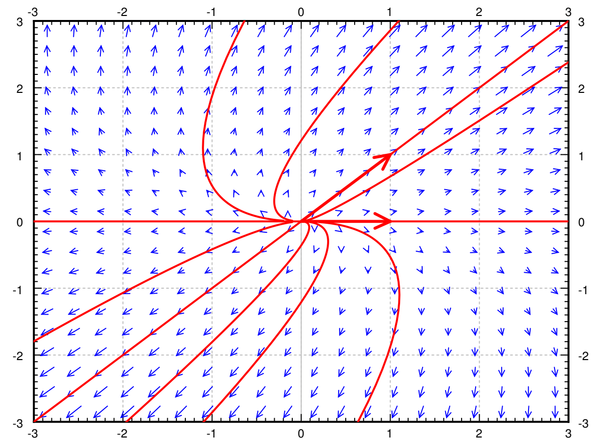


Figure 2.6: Example source vector field with eigenvectors and solutions.

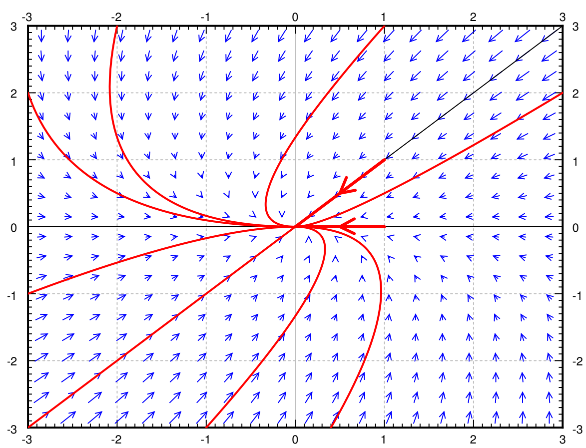


Figure 2.7: Example sink vector field with eigenvectors and solutions.

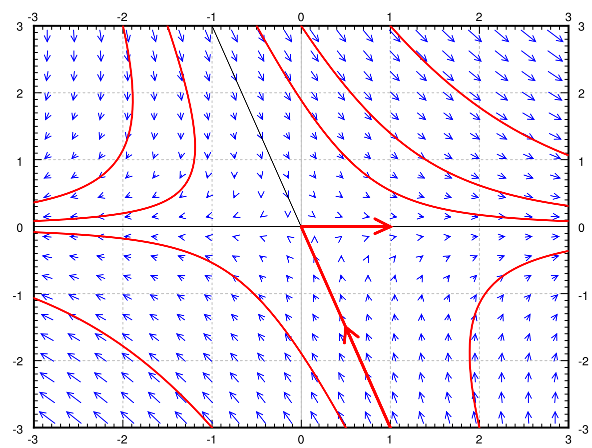


Figure 2.8: Example saddle vector field with eigenvectors and solutions.

Case 3. Suppose one eigenvalue is positive and one is negative. For example the matrix $\left[\begin{smallmatrix} 1 & 1 \\ 0 & -2 \end{smallmatrix}\right]$. The eigenvalues are 1 and -2 and corresponding eigenvectors are $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$. We reverse the arrows on one line (corresponding to the negative eigenvalue) and we obtain the picture in Figure 2.8. We call this picture a *saddle point*.

For the next three cases we will assume the eigenvalues are complex. In this case the eigenvectors are also complex and we cannot just plot them in the plane.

Case 4. Suppose the eigenvalues are purely imaginary, that is, $\pm ib$. For example, let $P = \left[\begin{smallmatrix} 0 & 1 \\ -4 & 0 \end{smallmatrix}\right]$. The eigenvalues are $\pm 2i$ and corresponding eigenvectors are $\left[\begin{smallmatrix} 1 \\ 2i \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ -2i \end{smallmatrix}\right]$. Consider the eigenvalue $2i$ and its eigenvector $\left[\begin{smallmatrix} 1 \\ 2i \end{smallmatrix}\right]$. The real and imaginary parts of $\vec{v}e^{2it}$

are

$$\operatorname{Re}\begin{bmatrix}1\\2i\end{bmatrix}e^{2it} = \begin{bmatrix}\cos(2t)\\-2\sin(2t)\end{bmatrix}, \qquad \operatorname{Im}\begin{bmatrix}1\\2i\end{bmatrix}e^{2it} = \begin{bmatrix}\sin(2t)\\2\cos(2t)\end{bmatrix}.$$

We can take any linear combination of them to get other solutions, which one we take depends on the initial conditions. Now note that the real part is a parametric equation for an ellipse. Same with the imaginary part and in fact any linear combination of the two. This is what happens in general when the eigenvalues are purely imaginary. So when the eigenvalues are purely imaginary, we get *ellipses* for the solutions. This type of picture is sometimes called a *center*. See Figure 2.9.

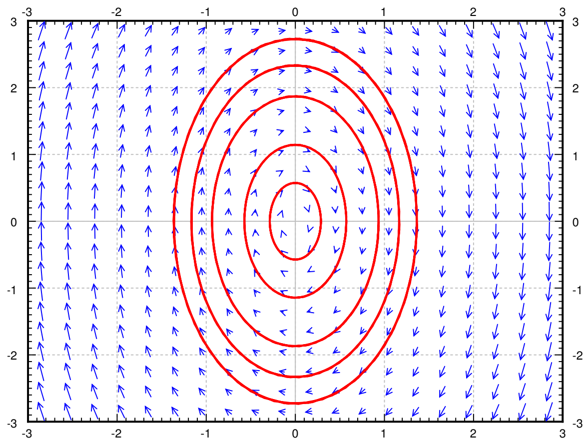


Figure 2.9: Example center vector field.

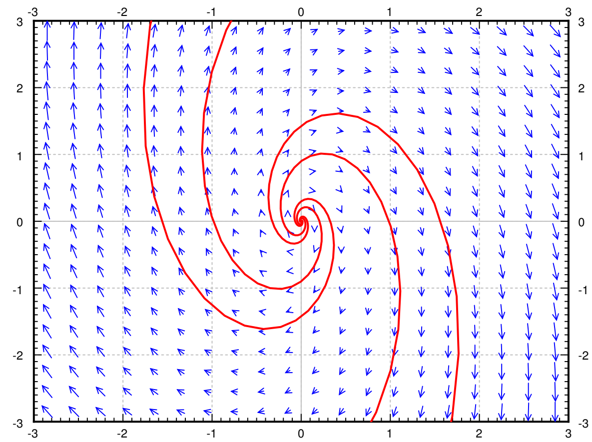


Figure 2.10: Example spiral source vector field.

Case 5. Now suppose the complex eigenvalues have a positive real part. That is, suppose the eigenvalues are $a \pm ib$ for some $a > 0$. For example, let $P = \left[\begin{smallmatrix}1 & 1\\-4 & 1\end{smallmatrix}\right]$. The eigenvalues turn out to be $1 \pm 2i$ and eigenvectors are $\left[\begin{smallmatrix}1\\2i\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\-2i\end{smallmatrix}\right]$. We take $1+2i$ and its eigenvector $\left[\begin{smallmatrix}1\\2i\end{smallmatrix}\right]$ and find the real and imaginary parts of $\vec{v}e^{(1+2i)t}$ are

$$\operatorname{Re}\begin{bmatrix}1\\2i\end{bmatrix}e^{(1+2i)t} = e^{t}\begin{bmatrix}\cos(2t)\\-2\sin(2t)\end{bmatrix}, \qquad \operatorname{Im}\begin{bmatrix}1\\2i\end{bmatrix}e^{(1+2i)t} = e^{t}\begin{bmatrix}\sin(2t)\\2\cos(2t)\end{bmatrix}.$$

Note the e^t in front of the solutions. The solutions grow in magnitude while spinning around the origin. Hence we get a *spiral source*. See Figure 2.10.

Case 6. Finally suppose the complex eigenvalues have a negative real part. That is, suppose the eigenvalues are $-a \pm ib$ for some $a > 0$. For example, let $P = \left[\begin{smallmatrix}-1 & -1\\4 & -1\end{smallmatrix}\right]$. The eigenvalues turn out to be $-1 \pm 2i$ and eigenvectors are $\left[\begin{smallmatrix}1\\-2i\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\2i\end{smallmatrix}\right]$. We take $-1-2i$ and its eigenvector $\left[\begin{smallmatrix}1\\2i\end{smallmatrix}\right]$ and find the real and imaginary parts of $\vec{v}e^{(-1-2i)t}$ are

$$\operatorname{Re}\begin{bmatrix}1\\2i\end{bmatrix}e^{(-1-2i)t} = e^{-t}\begin{bmatrix}\cos(2t)\\2\sin(2t)\end{bmatrix}, \qquad \operatorname{Im}\begin{bmatrix}1\\2i\end{bmatrix}e^{(-1-2i)t} = e^{-t}\begin{bmatrix}-\sin(2t)\\2\cos(2t)\end{bmatrix}.$$

Note the e^{-t} in front of the solutions. The solutions shrink in magnitude while spinning around the origin. Hence we get a *spiral sink*. See Figure 2.11.

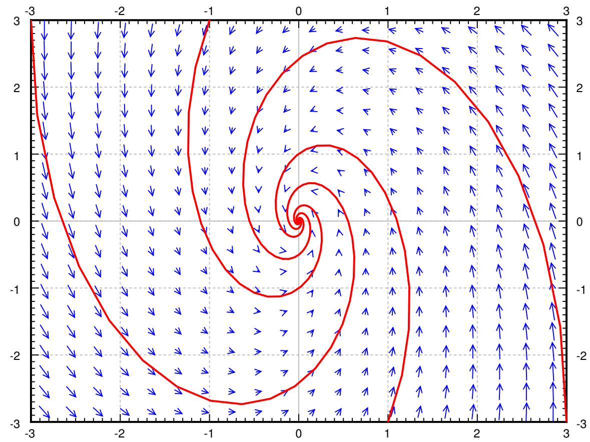


Figure 2.11: Example spiral sink vector field.

We summarize the behavior of linear homogeneous two-dimensional systems given by a nonsingular matrix in Table 2.1. Systems where one of the eigenvalues is zero (the matrix is singular) come up in practice from time to time, see Example 2.1.2 on page 73, and the pictures are somewhat different (simpler in a way). See the exercises.

Eigenvalues	Behavior
real and both positive	source / unstable node
real and both negative	sink / stable node
real and opposite signs	saddle
purely imaginary	center point / ellipses
complex with positive real part	spiral source
complex with negative real part	spiral sink

Table 2.1: Summary of behavior of linear homogeneous two-dimensional systems.

2.5.1 Exercises

***Exercise* 2.5.1:** *Take the equation $mx'' + cx' + kx = 0$, with $m > 0$, $c \geq 0$, $k > 0$ for the mass-spring system.*

a) *Convert this to a system of first order equations.*

b) *Classify for what m, c, k do you get which behavior.*

c) *Explain from physical intuition why you do not get all the different kinds of behavior here?*

Exercise* 2.5.2*: *What happens in the case when $P = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$? In this case the eigenvalue is repeated and there is only one independent eigenvector. What picture does this look like?*

Exercise* 2.5.3*: *What happens in the case when $P = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$? Does this look like any of the pictures we have drawn?*

Exercise* 2.5.4*: *Which behaviors are possible if P is diagonal, that is $P = \left[\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right]$? You can assume that a and b are not zero.*

Exercise* 2.5.5*: *Take the system from Example 2.1.2 on page 73, $x_1' = \frac{r}{V}(x_2 - x_1)$, $x_2' = \frac{r}{V}(x_1 - x_2)$. As we said, one of the eigenvalues is zero. What is the other eigenvalue, how does the picture look like and what happens when t goes to infinity.*

Exercise* 2.5.101*: *Describe the behavior of the following systems without solving:*

a) $x' = x + y, \quad y' = x - y.$

b) $x_1' = x_1 + x_2, \quad x_2' = 2x_2.$

c) $x_1' = -2x_2, \quad x_2' = 2x_1.$

d) $x' = x + 3y, \quad y' = -2x - 4y.$

e) $x' = x - 4y, \quad y' = -4x + y.$

Exercise* 2.5.102*: *Suppose that $\vec{x}\,' = A\vec{x}$ where A is a 2 by 2 matrix with eigenvalues $2 \pm i$. Describe the behavior.*

Exercise* 2.5.103*: *Take $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]' = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. Draw the vector field and describe the behavior. Is it one of the behaviors that we have seen before?*

2.6 Second order linear ODEs

Note: 1 lecture, reduction of order optional, first part of §3.1 in [EP], parts of §3.1 and §3.2 in [BD]

Consider the general *second order linear differential equation*

$$A(x)y'' + B(x)y' + C(x)y = F(x).$$

We usually divide through by $A(x)$ to get

$$y'' + p(x)y' + q(x)y = f(x), \tag{2.4}$$

where $p(x) = B(x)/A(x)$, $q(x) = C(x)/A(x)$, and $f(x) = F(x)/A(x)$. The word *linear* means that the equation contains no powers nor functions of y, y', and y''.

In the special case when $f(x) = 0$, we have a so-called *homogeneous* equation

$$y'' + p(x)y' + q(x)y = 0. \tag{2.5}$$

We have already seen some second order linear homogeneous equations.

$$y'' + k^2 y = 0 \qquad \text{Two solutions are:} \quad y_1 = \cos(kx), \quad y_2 = \sin(kx).$$
$$y'' - k^2 y = 0 \qquad \text{Two solutions are:} \quad y_1 = e^{kx}, \quad y_2 = e^{-kx}.$$

If we know two solutions of a linear homogeneous equation, we know many more of them.

Theorem 2.6.1 (Superposition)**.** *Suppose y_1 and y_2 are two solutions of the homogeneous equation* (2.5)*. Then*

$$y(x) = C_1 y_1(x) + C_2 y_2(x),$$

also solves (2.5) *for arbitrary constants C_1 and C_2.*

That is, we can add solutions together and multiply them by constants to obtain new and different solutions. We call the expression $C_1 y_1 + C_2 y_2$ a *linear combination* of y_1 and y_2. Let us prove this theorem; the proof is very enlightening and illustrates how linear equations work.

Proof: Let $y = C_1 y_1 + C_2 y_2$. Then

$$\begin{aligned} y'' + py' + qy &= (C_1 y_1 + C_2 y_2)'' + p(C_1 y_1 + C_2 y_2)' + q(C_1 y_1 + C_2 y_2) \\ &= C_1 y_1'' + C_2 y_2'' + C_1 p y_1' + C_2 p y_2' + C_1 q y_1 + C_2 q y_2 \\ &= C_1 (y_1'' + p y_1' + q y_1) + C_2 (y_2'' + p y_2' + q y_2) \\ &= C_1 \cdot 0 + C_2 \cdot 0 = 0. \qquad \square \end{aligned}$$

The proof becomes even simpler to state if we use the operator notation. An *operator* is an object that eats functions and spits out functions (kind of like what a function is, but a function eats numbers and spits out numbers). Define the operator L by

$$Ly = y'' + py' + qy.$$

The differential equation now becomes $Ly = 0$. The operator (and the equation) L being *linear* means that $L(C_1 y_1 + C_2 y_2) = C_1 L y_1 + C_2 L y_2$. It is almost as if we were "multiplying" by L. The proof above becomes

$$Ly = L(C_1 y_1 + C_2 y_2) = C_1 L y_1 + C_2 L y_2 = C_1 \cdot 0 + C_2 \cdot 0 = 0.$$

Two different solutions to the second equation $y'' - k^2 y = 0$ are $y_1 = \cosh(kx)$ and $y_2 = \sinh(kx)$. Recalling the definition of sinh and cosh, we note that these are solutions by superposition as they are linear combinations of the two exponential solutions: $\cosh(kx) = \frac{e^{kx}+e^{-kx}}{2} = (1/2)e^{kx} + (1/2)e^{-kx}$ and $\sinh(kx) = \frac{e^{kx}-e^{-kx}}{2} = (1/2)e^{kx} - (1/2)e^{-kx}$.

The functions sinh and cosh are sometimes more convenient to use than the exponential. Let us review some of their properties:

$$\cosh 0 = 1, \qquad \sinh 0 = 0,$$
$$\frac{d}{dx}\Big[\cosh x\Big] = \sinh x, \qquad \frac{d}{dx}\Big[\sinh x\Big] = \cosh x,$$
$$\cosh^2 x - \sinh^2 x = 1.$$

Exercise* 2.6.1*: *Derive these properties using the definitions of* sinh *and* cosh *in terms of exponentials.*

Linear equations have nice and simple answers to the existence and uniqueness question.

Theorem 2.6.2 (Existence and uniqueness)**.** *Suppose p, q, f are continuous functions on some interval I, a is a number in I, and b_0, b_1 are constants. Then the equation*

$$y'' + p(x)y' + q(x)y = f(x),$$

has exactly one solution $y(x)$ defined on the interval I satisfying the initial conditions

$$y(a) = b_0, \qquad y'(a) = b_1.$$

For example, the equation $y'' + k^2y = 0$ with $y(0) = b_0$ and $y'(0) = b_1$ has the solution

$$y(x) = b_0 \cos(kx) + \frac{b_1}{k} \sin(kx).$$

The equation $y'' - k^2y = 0$ with $y(0) = b_0$ and $y'(0) = b_1$ has the solution

$$y(x) = b_0 \cosh(kx) + \frac{b_1}{k} \sinh(kx).$$

Using cosh and sinh in this solution allows us to solve for the initial conditions in a cleaner way than if we have used the exponentials.

The initial conditions for a second order ODE consist of two equations. Common sense tells us that if we have two arbitrary constants and two equations, then we should be able to solve for the constants and find a solution to the differential equation satisfying the initial conditions.

Question: Suppose we find two different solutions y_1 and y_2 to the homogeneous equation (2.5). Can every solution be written (using superposition) in the form $y = C_1y_1 + C_2y_2$?

Answer is affirmative! Provided that y_1 and y_2 are different enough in the following sense. We say y_1 and y_2 are *linearly independent* if one is not a constant multiple of the other.

Theorem 2.6.3. *Let p, q be continuous functions. Let y_1 and y_2 be two linearly independent solutions to the homogeneous equation* (2.5)*. Then every other solution is of the form*

$$y = C_1y_1 + C_2y_2.$$

That is, $y = C_1y_1 + C_2y_2$ is the general solution.

For example, we found the solutions $y_1 = \sin x$ and $y_2 = \cos x$ for the equation $y'' + y = 0$. It is not hard to see that sine and cosine are not constant multiples of each other. Indeed, if $\sin x = A \cos x$ for some constant A, plugging in $x = 0$ would imply $A = 0$. But then $\sin x = 0$ for all x, which is preposterous. So y_1 and y_2 are linearly independent. Hence,

$$y = C_1 \cos x + C_2 \sin x$$

is the general solution to $y'' + y = 0$.

For two functions, checking linear independence is rather simple. Let us see another example. Consider $y'' - 2x^{-2}y = 0$. Then $y_1 = x^2$ and $y_2 = 1/x$ are solutions. To see that they are linearly independent, suppose one is a multiple of the other: $y_1 = Ay_2$, we just have to find out that A cannot be a constant. In this case we have $A = y_1/y_2 = x^3$, this most decidedly not a constant. So $y = C_1x^2 + C_2 1/x$ is the general solution.

If you have one solution to a second order linear homogeneous equation, then you can find another one. This is the *reduction of order method*. The idea is that if we somehow found y_1 as a solution of $y'' + p(x)y' + q(x)y = 0$, then we try a second solution of the form $y_2(x) = y_1(x)v(x)$. We just need to find v. We plug y_2 into the equation:

$$0 = y_2'' + p(x)y_2' + q(x)y_2 = \underbrace{y_1''v + 2y_1'v' + y_1v''}_{y_2''} + p(x)\underbrace{(y_1'v + y_1v')}_{y_2'} + q(x)\underbrace{y_1v}_{y_2}$$

$$= y_1v'' + (2y_1' + p(x)y_1)v' + \cancelto{0}{\big(y_1'' + p(x)y_1' + q(x)y_1\big)}v.$$

In other words, $y_1v'' + (2y_1' + p(x)y_1)v' = 0$. Using $w = v'$, we have the first order linear equation $y_1w' + (2y_1' + p(x)y_1)w = 0$. After solving this equation for w (integrating factor), we find v by antidifferentiating w. We then form y_2 by computing y_1v. For example, suppose we somehow know $y_1 = x$ is a solution to $y'' + x^{-1}y' - x^{-2}y = 0$. The equation for w is then $xw' + 3w = 0$. We find a solution, $w = Cx^{-3}$, and we find an antiderivative $v = \frac{-C}{2x^2}$. Hence $y_2 = y_1v = \frac{-C}{2x}$. Any C works and so $C = -2$ makes $y_2 = 1/x$. Thus, the general solution is $y = C_1x + C_2 1/x$.

Since we have a formula for the solution to the first order linear equation, we can write a formula for y_2:

$$y_2(x) = y_1(x)\int \frac{e^{-\int p(x)\,dx}}{\big(y_1(x)\big)^2}\,dx$$

However, it is much easier to remember that we just need to try $y_2(x) = y_1(x)v(x)$ and find $v(x)$ as we did above. Also, the technique works for higher order equations too: You get to reduce the order for each solution you find. So it is better to remember how to do it rather than a specific formula.

We will study the solution of nonhomogeneous equations in § **??**. We will first focus on finding general solutions to homogeneous equations.

2.6.1 Exercises

***Exercise* 2.6.2:** *Show that $y = e^x$ and $y = e^{2x}$ are linearly independent.*

***Exercise* 2.6.3:** *Take $y'' + 5y = 10x + 5$. Find (guess!) a solution.*

***Exercise* 2.6.4:** *Prove the superposition principle for nonhomogeneous equations. Suppose that y_1 is a solution to $Ly_1 = f(x)$ and y_2 is a solution to $Ly_2 = g(x)$ (same linear operator L). Show that $y = y_1 + y_2$ solves $Ly = f(x) + g(x)$.*

Exercise* 2.6.5*: *For the equation $x^2y'' - xy' = 0$, find two solutions, show that they are linearly independent and find the general solution. Hint: Try $y = x^r$.*

Equations of the form $ax^2y'' + bxy' + cy = 0$ are called *Euler's equations* or *Cauchy–Euler equations*. They are solved by trying $y = x^r$ and solving for r (assume $x \geq 0$ for simplicity).

Exercise* 2.6.6*: *Suppose that $(b - a)^2 - 4ac > 0$.*

a) Find a formula for the general solution of Euler's equation (see above) $ax^2y'' + bxy' + cy = 0$. Hint: Try $y = x^r$ and find a formula for r.

b) What happens when $(b - a)^2 - 4ac = 0$ or $(b - a)^2 - 4ac < 0$?

We will revisit the case when $(b - a)^2 - 4ac < 0$ later.

Exercise* 2.6.7*: *Same equation as in Exercise 2.6.6. Suppose $(b - a)^2 - 4ac = 0$. Find a formula for the general solution of $ax^2y'' + bxy' + cy = 0$. Hint: Try $y = x^r \ln x$ for the second solution.*

Exercise* 2.6.8** (reduction of order):*** *Suppose y_1 is a solution to $y'' + p(x)y' + q(x)y = 0$. By directly plugging into the equation, show that*

$$y_2(x) = y_1(x) \int \frac{e^{-\int p(x)\,dx}}{\big(y_1(x)\big)^2}\,dx$$

is also a solution.

Exercise* 2.6.9** (Chebyshev's equation of order 1):*** *Take $(1 - x^2)y'' - xy' + y = 0$.*

a) Show that $y = x$ is a solution.

b) Use reduction of order to find a second linearly independent solution.

c) Write down the general solution.

Exercise* 2.6.10** (Hermite's equation of order 2):*** *Take $y'' - 2xy' + 4y = 0$.*

a) Show that $y = 1 - 2x^2$ is a solution.

b) Use reduction of order to find a second linearly independent solution. (It's OK to leave a definite integral in the formula.)

c) Write down the general solution.

Exercise* 2.6.101*: *Are $\sin(x)$ and e^x linearly independent? Justify.*

Exercise* 2.6.102*: *Are e^x and e^{x+2} linearly independent? Justify.*

Exercise* 2.6.103*: *Guess a solution to $y'' + y' + y = 5$.*

Exercise* 2.6.104*: *Find the general solution to $xy'' + y' = 0$. Hint: It is a first order ODE in y'.*

Exercise* 2.6.105*: *Write down an equation (guess) for which we have the solutions e^x and e^{2x}. Hint: Try an equation of the form $y'' + Ay' + By = 0$ for constants A and B, plug in both e^x and e^{2x} and solve for A and B.*

2.7 Second order systems and applications

Note: more than 2 lectures, §5.4 in [EP], not in [BD]

2.7.1 Undamped mass-spring systems

While we did say that we will usually only look at first order systems, it is sometimes more convenient to study the system in the way it arises naturally. For example, suppose we have 3 masses connected by springs between two walls. We could pick any higher number, and the math would be essentially the same, but for simplicity we pick 3 right now. Let us also assume no friction, that is, the system is undamped. The masses are m_1, m_2, and m_3 and the spring constants are k_1, k_2, k_3, and k_4. Let x_1 be the displacement from rest position of the first mass, and x_2 and x_3 the displacement of the second and third mass. We make, as usual, positive values go right (as x_1 grows, the first mass is moving right). See Figure 2.12.

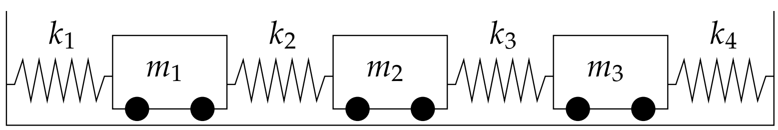


Figure 2.12: System of masses and springs.

This simple system turns up in unexpected places. For example, our world really consists of many small particles of matter interacting together. When we try the system above with many more masses, we obtain a good approximation to how an elastic material behaves. By somehow taking a limit of the number of masses going to infinity, we obtain the continuous one-dimensional wave equation (that we study in § **??**). But we digress.

Let us set up the equations for the three mass system. By Hooke's law, the force acting on the mass equals the spring compression times the spring constant. By Newton's second law, force is mass times acceleration. So if we sum the forces acting on each mass, put the right sign in front of each term, depending on the direction in which it is acting, and set this equal to mass times the acceleration, we end up with the desired system of equations.

$$\begin{aligned} m_1 x_1'' &= -k_1 x_1 + k_2(x_2 - x_1) &&= -(k_1 + k_2)x_1 + k_2 x_2, \\ m_2 x_2'' &= -k_2(x_2 - x_1) + k_3(x_3 - x_2) &&= k_2 x_1 - (k_2 + k_3)x_2 + k_3 x_3, \\ m_3 x_3'' &= -k_3(x_3 - x_2) - k_4 x_3 &&= k_3 x_2 - (k_3 + k_4)x_3. \end{aligned}$$

We define the matrices

$$M = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \qquad \text{and} \qquad K = \begin{bmatrix} -(k_1 + k_2) & k_2 & 0 \\ k_2 & -(k_2 + k_3) & k_3 \\ 0 & k_3 & -(k_3 + k_4) \end{bmatrix}.$$

We write the equation simply as

$$M\vec{x}'' = K\vec{x}.$$

At this point we could introduce 3 new variables and write out a system of 6 first order equations. We claim this simple setup is easier to handle as a second order system. We call $\vec{x}$ the *displacement vector*, M the *mass matrix*, and K the *stiffness matrix*.

Exercise* 2.7.1*: *Repeat this setup for 4 masses (find the matrices M and K). Do it for 5 masses. Can you find a prescription to do it for n masses?*

As with a single equation we want to "divide by M." This means computing the inverse of M. The masses are all nonzero and M is a diagonal matrix, so computing the inverse is easy:

$$M^{-1} = \begin{bmatrix} \frac{1}{m_1} & 0 & 0 \\ 0 & \frac{1}{m_2} & 0 \\ 0 & 0 & \frac{1}{m_3} \end{bmatrix}.$$

This fact follows readily by how we multiply diagonal matrices. As an exercise, you should verify that $MM^{-1} = M^{-1}M = I$.

Let $A = M^{-1}K$. We look at the system $\vec{x}'' = M^{-1}K\vec{x}$, or

$$\vec{x}'' = A\vec{x}.$$

Many real world systems can be modeled by this equation. For simplicity, we will only talk about the given masses-and-springs problem. We try a solution of the form

$$\vec{x} = \vec{v}e^{\alpha t}.$$

We compute that for this guess, $\vec{x}'' = \alpha^2\vec{v}e^{\alpha t}$. We plug our guess into the equation and get

$$\alpha^2\vec{v}e^{\alpha t} = A\vec{v}e^{\alpha t}.$$

We divide by $e^{\alpha t}$ to arrive at $\alpha^2\vec{v} = A\vec{v}$. Hence if α^2 is an eigenvalue of A and $\vec{v}$ is a corresponding eigenvector, we have found a solution.

In our example, and in other common applications, A has only real negative eigenvalues (and possibly a zero eigenvalue). So we study only this case. When an eigenvalue λ is negative, it means that $\alpha^2 = \lambda$ is negative. Hence there is some real number ω such that $-\omega^2 = \lambda$. Then $\alpha = \pm i\omega$. The solution we guessed was

$$\vec{x} = \vec{v}\,\big(\cos(\omega t) + i\sin(\omega t)\big).$$

By taking the real and imaginary parts (note that $\vec{v}$ is real), we find that $\vec{v}\cos(\omega t)$ and $\vec{v}\sin(\omega t)$ are linearly independent solutions.

If an eigenvalue is zero, it turns out that both $\vec{v}$ and $\vec{v}t$ are solutions, where $\vec{v}$ is an eigenvector corresponding to the eigenvalue 0.

***Exercise* 2.7.2:** *Show that if A has a zero eigenvalue and $\vec{v}$ is a corresponding eigenvector, then $\vec{x} = \vec{v}(a + bt)$ is a solution of $\vec{x}'' = A\vec{x}$ for arbitrary constants a and b.*

Theorem 2.7.1. *Let A be a real $n \times n$ matrix with n distinct real negative (or zero) eigenvalues we denote by $-\omega_1^2 > -\omega_2^2 > \cdots > -\omega_n^2$, and corresponding eigenvectors by $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. If A is invertible (that is, if $\omega_1 > 0$), then*

$$\boxed{\vec{x}(t) = \sum_{i=1}^{n} \vec{v}_i \big(a_i \cos(\omega_i t) + b_i \sin(\omega_i t)\big),}$$

is the general solution of

$$\vec{x}'' = A\vec{x},$$

for some arbitrary constants a_i and b_i. If A has a zero eigenvalue, that is $\omega_1 = 0$, and all other eigenvalues are distinct and negative, then the general solution can be written as

$$\boxed{\vec{x}(t) = \vec{v}_1(a_1 + b_1 t) + \sum_{i=2}^{n} \vec{v}_i \big(a_i \cos(\omega_i t) + b_i \sin(\omega_i t)\big).}$$

We use this solution and the setup from the introduction of this section even when some of the masses and springs are missing. For example, when there are only 2 masses and only 2 springs, simply take only the equations for the two masses and set all the spring constants for the springs that are missing to zero.

2.7.2 Examples

Example 2.7.1: Consider the setup in Figure 2.13, with $m_1 = 2\,\text{kg}$, $m_2 = 1\,\text{kg}$, $k_1 = 4\,\text{N/m}$, and $k_2 = 2\,\text{N/m}$.

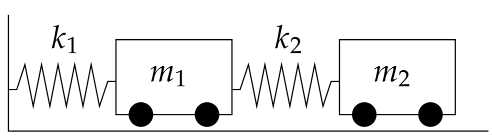


Figure 2.13: System of masses and springs.

The equations we write down are

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \vec{x}'' = \begin{bmatrix} -(4+2) & 2 \\ 2 & -2 \end{bmatrix} \vec{x},$$

or

$$\vec{x}'' = \begin{bmatrix} -3 & 1 \\ 2 & -2 \end{bmatrix} \vec{x}.$$

We find the eigenvalues of A to be $\lambda = -1, -4$ (exercise). We find corresponding eigenvectors to be $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$ respectively (exercise).

We check the theorem and note that $\omega_1 = 1$ and $\omega_2 = 2$. Hence the general solution is

$$\vec{x} = \begin{bmatrix}1\\2\end{bmatrix}\big(a_1\cos(t) + b_1\sin(t)\big) + \begin{bmatrix}1\\-1\end{bmatrix}\big(a_2\cos(2t) + b_2\sin(2t)\big).$$

The two terms in the solution represent the two so-called *natural* or *normal modes of oscillation*. And the two (angular) frequencies are the *natural frequencies*. The first natural frequency is 1, and second natural frequency is 2. The two modes are plotted in Figure 2.14.

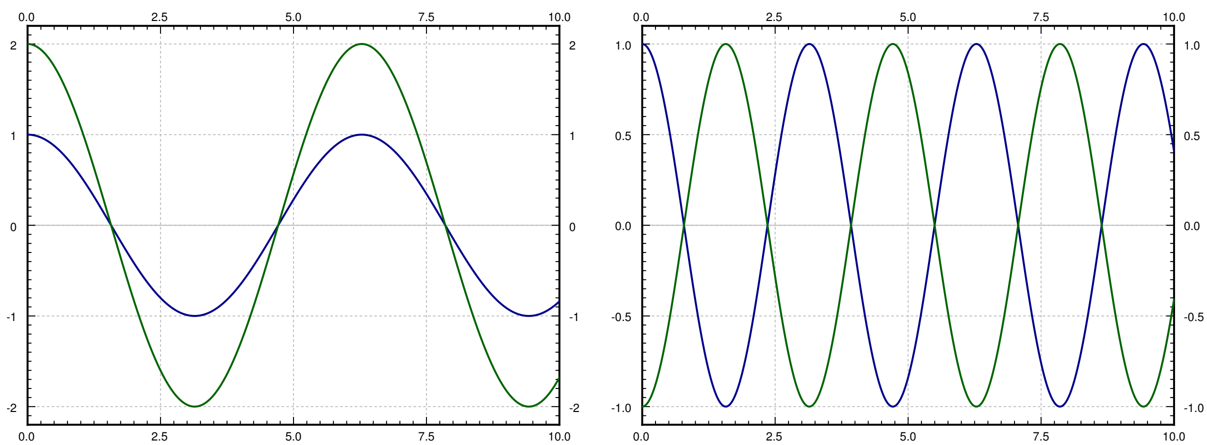


Figure 2.14: The two modes of the mass-spring system. In the left plot the masses are moving in unison and in the right plot are masses moving in the opposite direction.

Let us write the solution as

$$\vec{x} = \begin{bmatrix}1\\2\end{bmatrix} c_1\cos(t - \alpha_1) + \begin{bmatrix}1\\-1\end{bmatrix} c_2\cos(2t - \alpha_2).$$

The first term,

$$\begin{bmatrix}1\\2\end{bmatrix} c_1\cos(t - \alpha_1) = \begin{bmatrix}c_1\cos(t - \alpha_1)\\2c_1\cos(t - \alpha_1)\end{bmatrix},$$

corresponds to the mode where the masses move synchronously in the same direction.

The second term,

$$\begin{bmatrix}1\\-1\end{bmatrix} c_2\cos(2t - \alpha_2) = \begin{bmatrix}c_2\cos(2t - \alpha_2)\\-c_2\cos(2t - \alpha_2)\end{bmatrix},$$

corresponds to the mode where the masses move synchronously but in opposite directions.

The general solution is a combination of the two modes. That is, the initial conditions determine the amplitude and phase shift of each mode. As an example, suppose we have initial conditions

$$\vec{x}(0) = \begin{bmatrix}1\\-1\end{bmatrix}, \qquad \vec{x}'(0) = \begin{bmatrix}0\\6\end{bmatrix}.$$

We use the a_j, b_j constants to solve for initial conditions. First

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \vec{x}(0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} a_1 + \begin{bmatrix} 1 \\ -1 \end{bmatrix} a_2 = \begin{bmatrix} a_1 + a_2 \\ 2a_1 - a_2 \end{bmatrix}.$$

We solve (exercise) to find $a_1 = 0$, $a_2 = 1$. To find the b_1 and b_2, we differentiate first:

$$\vec{x}' = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \big(-a_1 \sin(t) + b_1 \cos(t)\big) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \big(-2a_2 \sin(2t) + 2b_2 \cos(2t)\big).$$

Now we solve:

$$\begin{bmatrix} 0 \\ 6 \end{bmatrix} = \vec{x}'(0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} b_1 + \begin{bmatrix} 1 \\ -1 \end{bmatrix} 2b_2 = \begin{bmatrix} b_1 + 2b_2 \\ 2b_1 - 2b_2 \end{bmatrix}.$$

Again solve (exercise) to find $b_1 = 2$, $b_2 = -1$. So our solution is

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} 2\sin(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \big(\cos(2t) - \sin(2t)\big) = \begin{bmatrix} 2\sin(t) + \cos(2t) - \sin(2t) \\ 4\sin(t) - \cos(2t) + \sin(2t) \end{bmatrix}.$$

The graphs of the two displacements, x_1 and x_2 of the two carts is in Figure 2.15.

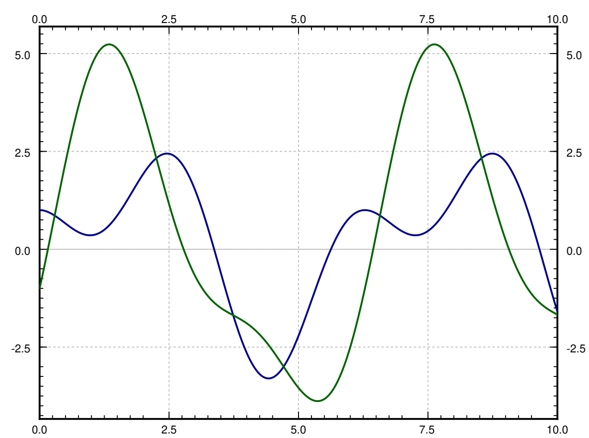


Figure 2.15: Superposition of the two modes given the initial conditions.

Example 2.7.2: We have two toy rail cars. Car 1 of mass 2 kg is traveling at 3 m/s towards the second rail car of mass 1 kg. There is a bumper on the second rail car that engages at the moment the cars hit (it connects to two cars) and does not let go. The bumper acts like a spring of spring constant $k = 2$ N/m. The second car is 10 meters from a wall. See Figure 2.16 on the next page.

We want to ask several questions. At what time after the cars link does impact with the wall happen? What is the speed of car 2 when it hits the wall?

OK, let us first set the system up. Let $t = 0$ be the time when the two cars link up. Let x_1 be the displacement of the first car from the position at $t = 0$, and let x_2 be the displacement

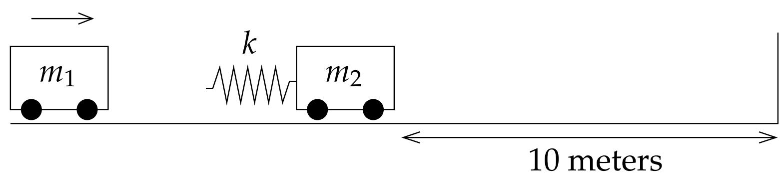


Figure 2.16: The crash of two rail cars.

of the second car from its original location. Then the time when $x_2(t) = 10$ is exactly the time when impact with wall occurs. For this t, $x_2'(t)$ is the speed at impact. This system acts just like the system of the previous example but without k_1. Hence the equation is

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \vec{x}'' = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \vec{x},$$

or

$$\vec{x}'' = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \vec{x}.$$

We compute the eigenvalues of A. It is not hard to see that the eigenvalues are 0 and -3 (exercise). Furthermore, eigenvectors are $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right]$ respectively (exercise). Then $\omega_1 = 0$, $\omega_2 = \sqrt{3}$, and by the second part of the theorem the general solution is

$$\begin{aligned} \vec{x} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} (a_1 + b_1 t) + \begin{bmatrix} 1 \\ -2 \end{bmatrix} \left(a_2 \cos(\sqrt{3}\, t) + b_2 \sin(\sqrt{3}\, t) \right) \\ &= \begin{bmatrix} a_1 + b_1 t + a_2 \cos(\sqrt{3}\, t) + b_2 \sin(\sqrt{3}\, t) \\ a_1 + b_1 t - 2a_2 \cos(\sqrt{3}\, t) - 2b_2 \sin(\sqrt{3}\, t) \end{bmatrix}. \end{aligned}$$

We now apply the initial conditions. First the cars start at position 0 so $x_1(0) = 0$ and $x_2(0) = 0$. The first car is traveling at $3\,{}^{\mathrm{m}}\!/\!{}_{\mathrm{s}}$, so $x_1'(0) = 3$ and the second car starts at rest, so $x_2'(0) = 0$. The first conditions says

$$\vec{0} = \vec{x}(0) = \begin{bmatrix} a_1 + a_2 \\ a_1 - 2a_2 \end{bmatrix}.$$

It is not hard to see that $a_1 = a_2 = 0$. We set $a_1 = 0$ and $a_2 = 0$ in $\vec{x}(t)$ and differentiate to get

$$\vec{x}'(t) = \begin{bmatrix} b_1 + \sqrt{3}\, b_2 \cos(\sqrt{3}\, t) \\ b_1 - 2\sqrt{3}\, b_2 \cos(\sqrt{3}\, t) \end{bmatrix}.$$

So

$$\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \vec{x}'(0) = \begin{bmatrix} b_1 + \sqrt{3}\, b_2 \\ b_1 - 2\sqrt{3}\, b_2 \end{bmatrix}.$$

Solving these two equations we find $b_1 = 2$ and $b_2 = \frac{1}{\sqrt{3}}$. Hence the position of our cars is (until the impact with the wall)

$$\vec{x} = \begin{bmatrix} 2t + \frac{1}{\sqrt{3}} \sin(\sqrt{3}\,t) \\ 2t - \frac{2}{\sqrt{3}} \sin(\sqrt{3}\,t) \end{bmatrix}.$$

Note how the presence of the zero eigenvalue resulted in a term containing t. This means that the cars will be traveling in the positive direction as time grows, which is what we expect.

What we are really interested in is the second expression, the one for x_2. We have $x_2(t) = 2t - \frac{2}{\sqrt{3}} \sin(\sqrt{3}\,t)$. See Figure 2.17 for the plot of x_2 versus time.

Just from the graph we can see that time of impact will be a little more than 5 seconds from time zero. For this we have to solve the equation $10 = x_2(t) = 2t - \frac{2}{\sqrt{3}} \sin(\sqrt{3}\,t)$. Using a computer (or even a graphing calculator) we find that $t_{\text{impact}} \approx 5.22$ seconds.

The speed of the second car is $x_2' = 2 - 2\cos(\sqrt{3}\,t)$. At the time of impact (5.22 seconds from $t = 0$) we get $x_2'(t_{\text{impact}}) \approx 3.85$. The maximum speed is the maximum of $2 - 2\cos(\sqrt{3}\,t)$, which is 4. We are traveling at almost the maximum speed when we hit the wall.

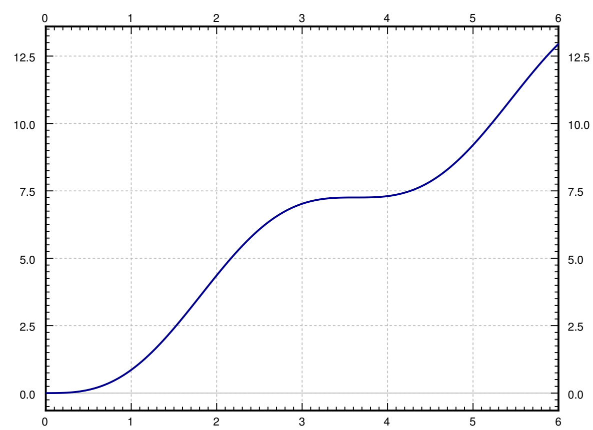


Figure 2.17: Position of the second car in time (ignoring the wall).

Suppose that Bob is a tiny person sitting on car 2. Bob has a Martini in his hand and would like not to spill it. Let us suppose Bob would not spill his Martini when the first car links up with car 2, but if car 2 hits the wall at any speed greater than zero, Bob will spill his drink. Suppose Bob can move car 2 a few meters towards or away from the wall (he cannot go all the way to the wall, nor can he get out of the way of the first car). Is there a "safe" distance for him to be at? A distance such that the impact with the wall is at zero speed?

The answer is yes. On Figure 2.17, note the "plateau" between $t = 3$ and $t = 4$. There is a point where the speed is zero. To find it we solve $x_2'(t) = 0$. This is when $\cos(\sqrt{3}\,t) = 1$ or in other words when $t = \frac{2\pi}{\sqrt{3}}, \frac{4\pi}{\sqrt{3}}, \ldots$ and so on. We plug in the first value to obtain $x_2\left(\frac{2\pi}{\sqrt{3}}\right) = \frac{4\pi}{\sqrt{3}} \approx 7.26$. So a "safe" distance is about 7 and a quarter meters from the wall.

Alternatively Bob could move away from the wall towards the incoming car 2, where another safe distance is $x_2\left(\frac{4\pi}{\sqrt{3}}\right) = \frac{8\pi}{\sqrt{3}} \approx 14.51$ and so on. We can use all the different t such that $x_2'(t) = 0$. Of course $t = 0$ is also a solution, corresponding to $x_2 = 0$, but that means standing right at the wall.

2.7.3 Forced oscillations

Finally we move to forced oscillations. Suppose that now our system is

$$\vec{x}'' = A\vec{x} + \vec{F}\cos(\omega t). \tag{2.6}$$

That is, we are adding periodic forcing to the system in the direction of the vector $\vec{F}$.

As before, this system just requires us to find one particular solution $\vec{x}_p$, add it to the general solution of the associated homogeneous system $\vec{x}_c$, and we will have the general solution to (2.6). Let us suppose that ω is not one of the natural frequencies of $\vec{x}'' = A\vec{x}$, then we can guess

$$\vec{x}_p = \vec{c}\cos(\omega t),$$

where $\vec{c}$ is an unknown constant vector. Note that we do not need to use sine since there are only second derivatives. We solve for $\vec{c}$ to find $\vec{x}_p$. This is really just the method of *undetermined coefficients* for systems. Let us differentiate $\vec{x}_p$ twice to get

$$\vec{x}_p'' = -\omega^2\vec{c}\cos(\omega t).$$

Plug $\vec{x}_p$ and $\vec{x}_p''$ into equation (2.6):

$$\overbrace{-\omega^2\vec{c}\cos(\omega t)}^{\vec{x}_p''} = \overbrace{A\vec{c}\cos(\omega t)}^{A\vec{x}_p} + \vec{F}\cos(\omega t).$$

We cancel out the cosine and rearrange the equation to obtain

$$(A + \omega^2 I)\vec{c} = -\vec{F}.$$

So

$$\vec{c} = (A + \omega^2 I)^{-1}(-\vec{F}).$$

Of course this is possible only if $(A + \omega^2 I) = \big(A - (-\omega^2)I\big)$ is invertible. That matrix is invertible if and only if $-\omega^2$ is not an eigenvalue of A. That is true if and only if ω is not a natural frequency of the system.

We simplified things a little bit. If we wish to have the forcing term to be in the units of force, say Newtons, then we must write

$$M\vec{x}'' = K\vec{x} + \vec{G}\cos(\omega t).$$

If we then write things in terms of $A = M^{-1}K$, we have

$$\vec{x}'' = M^{-1}K\vec{x} + M^{-1}\vec{G}\cos(\omega t) \qquad \text{or} \qquad \vec{x}'' = A\vec{x} + \vec{F}\cos(\omega t),$$

where $\vec{F} = M^{-1}\vec{G}$.

Example 2.7.3: Let us take the example in Figure 2.13 on page 110 with the same parameters as before: $m_1 = 2$, $m_2 = 1$, $k_1 = 4$, and $k_2 = 2$. Now suppose that there is a force $2\cos(3t)$ acting on the second cart.

The equation is

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \vec{x}'' = \begin{bmatrix} -(4+2) & 2 \\ 2 & -2 \end{bmatrix} \vec{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cos(3t) \qquad \text{or} \qquad \vec{x}'' = \begin{bmatrix} -3 & 1 \\ 2 & -2 \end{bmatrix} \vec{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cos(3t).$$

We solved the associated homogeneous equation before and found the complementary solution to be

$$\vec{x}_c = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \big(a_1 \cos(t) + b_1 \sin(t)\big) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \big(a_2 \cos(2t) + b_2 \sin(2t)\big).$$

The natural frequencies are 1 and 2. As 3 is not a natural frequency, we try $\vec{c} \cos(3t)$. We invert $(A + 3^2 I)$:

$$\left(\begin{bmatrix} -3 & 1 \\ 2 & -2 \end{bmatrix} + 3^2 I \right)^{-1} = \begin{bmatrix} 6 & 1 \\ 2 & 7 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{7}{40} & \frac{-1}{40} \\ \frac{-1}{20} & \frac{3}{20} \end{bmatrix}.$$

Hence,

$$\vec{c} = (A + \omega^2 I)^{-1} (-\vec{F}) = \begin{bmatrix} \frac{7}{40} & \frac{-1}{40} \\ \frac{-1}{20} & \frac{3}{20} \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{20} \\ \frac{-3}{10} \end{bmatrix}.$$

Combining with the general solution of the associated homogeneous problem, we get that the general solution to $\vec{x}'' = A\vec{x} + \vec{F} \cos(\omega t)$ is

$$\vec{x} = \vec{x}_c + \vec{x}_p = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \big(a_1 \cos(t) + b_1 \sin(t)\big) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \big(a_2 \cos(2t) + b_2 \sin(2t)\big) + \begin{bmatrix} \frac{1}{20} \\ \frac{-3}{10} \end{bmatrix} \cos(3t).$$

We would then solve for the constants a_1, a_2, b_1, and b_2 using any given initial conditions.

Note that given force $\vec{f}$, we write the equation as $M\vec{x}'' = K\vec{x} + \vec{f}$ to get the units right. Then we write $\vec{x}'' = M^{-1}K\vec{x} + M^{-1}\vec{f}$. The term $\vec{g} = M^{-1}\vec{f}$ in $\vec{x}'' = A\vec{x} + \vec{g}$ is in units of force per unit mass.

If ω is a natural frequency of the system, *resonance* may occur, because we will have to try a particular solution of the form

$$\vec{x}_p = \vec{c}\, t \sin(\omega t) + \vec{d} \cos(\omega t).$$

That is assuming that the eigenvalues of the coefficient matrix are distinct. Next, note that the amplitude of this solution grows without bound as t grows.

2.7.4 Exercises

***Exercise* 2.7.3:** *Find a particular solution to*

$$\vec{x}'' = \begin{bmatrix} -3 & 1 \\ 2 & -2 \end{bmatrix} \vec{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cos(2t).$$

Exercise* 2.7.4** (challenging):*** *Let us take the example in Figure 2.13 on page 110 with the same parameters as before: $m_1 = 2$, $k_1 = 4$, and $k_2 = 2$, except for m_2, which is unknown. Suppose that there is a force* $\cos(5t)$ *acting on the first mass. Find an m_2 such that there exists a particular solution where the first mass does not move.*

Note: This idea is called dynamic damping*. In practice there will be a small amount of damping and so any transient solution will disappear and after long enough time, the first mass will always come to a stop.*

Exercise* 2.7.5*: *Let us take the Example 2.7.2 on page 112, but that at time of impact, car 2 is moving to the left at the speed of 3 m/s.*

a) *Find the behavior of the system after linkup.*
b) *Will the second car hit the wall, or will it be moving away from the wall as time goes on?*
c) *At what speed would the first car have to be traveling for the system to essentially stay in place after linkup?*

Exercise* 2.7.6*: *Let us take the example in Figure 2.13 on page 110 with parameters $m_1 = m_2 = 1$, $k_1 = k_2 = 1$. Does there exist a set of initial conditions for which the first cart moves but the second cart does not? If so, find those conditions. If not, argue why not.*

Exercise* 2.7.101*: *Find the general solution to* $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \vec{x}\,'' = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -4 & 0 \\ 0 & 6 & -3 \end{bmatrix} \vec{x} + \begin{bmatrix} \cos(2t) \\ 0 \\ 0 \end{bmatrix}$.

Exercise* 2.7.102*: *Suppose there are three carts of equal mass m and connected by two springs of constant k (and no connections to walls). Set up the system and find its general solution.*

Exercise* 2.7.103*: *Suppose a cart of mass 2 kg is attached by a spring of constant $k = 1$ to a cart of mass 3 kg, which is attached to the wall by a spring also of constant $k = 1$. Suppose that the initial position of the first cart is 1 meter in the positive direction from the rest position, and the second mass starts at the rest position. The masses are not moving and are let go. Find the position of the second mass as a function of time.*

2.8 Multiple eigenvalues

Note: 1 or 1.5 lectures, §5.5 in [EP], §7.8 in [BD]

It may happen that a matrix A has some "repeated" eigenvalues. That is, the characteristic equation $\det(A - \lambda I) = 0$ may have repeated roots. This is actually unlikely to happen for a random matrix. If we take a small perturbation of A (we change the entries of A slightly), we get a matrix with distinct eigenvalues. As any system we want to solve in practice is an approximation to reality anyway, it is not absolutely indispensable to know how to solve these corner cases. On the other hand, these cases do come up in applications from time to time. Furthermore, if we have distinct but very close eigenvalues, the behavior is similar to that of repeated eigenvalues, and so understanding that case will give us insight into what is going on.

2.8.1 Geometric multiplicity

Take the diagonal matrix

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}.$$

A has an eigenvalue 3 of multiplicity 2. We call the multiplicity of the eigenvalue in the characteristic equation the *algebraic multiplicity*. In this case, there also exist 2 linearly independent eigenvectors, $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ corresponding to the eigenvalue 3. This means that the so-called *geometric multiplicity* of this eigenvalue is also 2.

In all the theorems where we required a matrix to have n distinct eigenvalues, we only really needed to have n linearly independent eigenvectors. For example, $\vec{x}\,' = A\vec{x}$ has the general solution

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{3t}.$$

Let us restate the theorem about real eigenvalues. In the following theorem we will repeat eigenvalues according to (algebraic) multiplicity. So for the matrix A above, we would say that it has eigenvalues 3 and 3.

Theorem 2.8.1. *Suppose the $n \times n$ matrix P has n real eigenvalues (not necessarily distinct), λ_1, $\lambda_2, \ldots, \lambda_n$, and there are n linearly independent corresponding eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. Then the general solution to $\vec{x}\,' = P\vec{x}$ can be written as*

$$\vec{x} = c_1 \vec{v}_1 e^{\lambda_1 t} + c_2 \vec{v}_2 e^{\lambda_2 t} + \cdots + c_n \vec{v}_n e^{\lambda_n t}.$$

The *geometric multiplicity* of an eigenvalue of algebraic multiplicity n is equal to the number of corresponding linearly independent eigenvectors. The geometric multiplicity is always less than or equal to the algebraic multiplicity. The theorem handles the case when these two multiplicities are equal for all eigenvalues. If for an eigenvalue the geometric multiplicity is equal to the algebraic multiplicity, then we say the eigenvalue is *complete*.

In other words, the hypothesis of the theorem could be stated as saying that if all the eigenvalues of P are complete, then there are n linearly independent eigenvectors and thus we have the given general solution.

If the geometric multiplicity of an eigenvalue is 2 or greater, then the set of linearly independent eigenvectors is not unique up to multiples as it was before. For example, for the diagonal matrix $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ we could also pick eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, or in fact any pair of two linearly independent vectors. The number of linearly independent eigenvectors corresponding to λ is the number of free variables we obtain when solving $A\vec{v} = \lambda\vec{v}$. We pick specific values for those free variables to obtain eigenvectors. If you pick different values, you may get different eigenvectors.

2.8.2 Defective eigenvalues

If an $n \times n$ matrix has less than n linearly independent eigenvectors, it is said to be *deficient*. Then there is at least one eigenvalue with an algebraic multiplicity that is higher than its geometric multiplicity. We call this eigenvalue *defective* and the difference between the two multiplicities we call the *defect*.

Example 2.8.1: The matrix

$$\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$$

has an eigenvalue 3 of algebraic multiplicity 2. Let us try to compute eigenvectors.

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \vec{0}.$$

We must have that $v_2 = 0$. Hence any eigenvector is of the form $\begin{bmatrix} v_1 \\ 0 \end{bmatrix}$. Any two such vectors are linearly dependent, and hence the geometric multiplicity of the eigenvalue is 1. Therefore, the defect is 1, and we can no longer apply the eigenvalue method directly to a system of ODEs with such a coefficient matrix.

Roughly, the key observation is that if λ is an eigenvalue of A of algebraic multiplicity m, then we can find certain m linearly independent vectors solving $(A - \lambda I)^k \vec{v} = \vec{0}$ for various powers k. We will call these *generalized eigenvectors*.

Let us continue with the example $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ and the equation $\vec{x}' = A\vec{x}$. We found an eigenvalue $\lambda = 3$ of (algebraic) multiplicity 2 and defect 1. We found one eigenvector $\vec{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We have one solution

$$\vec{x}_1 = \vec{v}e^{3t} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t}.$$

We are now stuck, we get no other solutions from standard eigenvectors. But we need two linearly independent solutions to find the general solution of the equation.

Let us try (in the spirit of repeated roots of the characteristic equation for a single equation) another solution of the form

$$\vec{x}_2 = (\vec{v}_2 + \vec{v}_1 t)\, e^{3t}.$$

We differentiate to get

$$\vec{x}_2' = \vec{v}_1 e^{3t} + 3(\vec{v}_2 + \vec{v}_1 t)\, e^{3t} = (3\vec{v}_2 + \vec{v}_1)\, e^{3t} + 3\vec{v}_1 t e^{3t}.$$

As we are assuming that $\vec{x}_2$ is a solution, $\vec{x}_2'$ must equal $A\vec{x}_2$. So we compute $A\vec{x}_2$:

$$A\vec{x}_2 = A(\vec{v}_2 + \vec{v}_1 t)\, e^{3t} = A\vec{v}_2 e^{3t} + A\vec{v}_1 t e^{3t}.$$

By looking at the coefficients of e^{3t} and te^{3t}, we see $3\vec{v}_2 + \vec{v}_1 = A\vec{v}_2$ and $3\vec{v}_1 = A\vec{v}_1$. This means that

$$(A - 3I)\vec{v}_2 = \vec{v}_1, \qquad \text{and} \qquad (A - 3I)\vec{v}_1 = \vec{0}.$$

Therefore, $\vec{x}_2$ is a solution if these two equations are satisfied. The second equation is satisfied if $\vec{v}_1$ is an eigenvector, and we found the eigenvector above, so let $\vec{v}_1 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. So, if we can find a $\vec{v}_2$ that solves $(A - 3I)\vec{v}_2 = \vec{v}_1$, then we are done. This is just a bunch of linear equations to solve and we are by now very good at that. Let us solve $(A - 3I)\vec{v}_2 = \vec{v}_1$. Write

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

By inspection we see that letting $a = 0$ (a could be anything in fact) and $b = 1$ does the job. Hence we can take $\vec{v}_2 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. Our general solution to $\vec{x}\,' = A\vec{x}$ is

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t} + c_2 \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} t \right) e^{3t} = \begin{bmatrix} c_1 e^{3t} + c_2 t e^{3t} \\ c_2 e^{3t} \end{bmatrix}.$$

Let us check that we really do have the solution. First $x_1' = c_1 3 e^{3t} + c_2 e^{3t} + 3c_2 t e^{3t} = 3x_1 + x_2$. Good. Now $x_2' = 3c_2 e^{3t} = 3x_2$. Good.

In the example, if we plug $(A - 3I)\vec{v}_2 = \vec{v}_1$ into $(A - 3I)\vec{v}_1 = \vec{0}$ we find

$$(A - 3I)(A - 3I)\vec{v}_2 = \vec{0}, \qquad \text{or} \qquad (A - 3I)^2 \vec{v}_2 = \vec{0}.$$

Furthermore, if $(A - 3I)\vec{w} \neq \vec{0}$, then $(A - 3I)\vec{w}$ is an eigenvector, a multiple of $\vec{v}_1$. In this 2×2 case $(A - 3I)^2$ is just the zero matrix (exercise). So any vector $\vec{w}$ solves $(A - 3I)^2\vec{w} = \vec{0}$ and we just need a $\vec{w}$ such that $(A - 3I)\vec{w} \neq \vec{0}$. Then we could use $\vec{w}$ for $\vec{v}_2$, and $(A - 3I)\vec{w}$ for $\vec{v}_1$.

Note that the system $\vec{x}\,' = A\vec{x}$ has a simpler solution since A is a so-called *upper triangular matrix*, that is every entry below the diagonal is zero. In particular, the equation for x_2 does not depend on x_1. Mind you, not every defective matrix is triangular.

***Exercise* 2.8.1:** *Solve* $\vec{x}' = \left[\begin{smallmatrix} 3 & 1 \\ 0 & 3 \end{smallmatrix}\right] \vec{x}$ *by first solving for* x_2 *and then for* x_1 *independently. Check that you got the same solution as we did above.*

Let us describe the general algorithm. Suppose that λ is an eigenvalue of multiplicity 2, defect 1. First find an eigenvector $\vec{v}_1$ of λ. That is, $\vec{v}_1$ solves $(A - \lambda I)\vec{v}_1 = \vec{0}$. Then, find a vector $\vec{v}_2$ such that

$$(A - \lambda I)\vec{v}_2 = \vec{v}_1.$$

This gives us two linearly independent solutions

$$\vec{x}_1 = \vec{v}_1 e^{\lambda t},$$
$$\vec{x}_2 = \left(\vec{v}_2 + \vec{v}_1 t\right) e^{\lambda t}.$$

Example 2.8.2: Consider the system

$$\vec{x}' = \begin{bmatrix} 2 & -5 & 0 \\ 0 & 2 & 0 \\ -1 & 4 & 1 \end{bmatrix} \vec{x}.$$

Compute the eigenvalues,

$$0 = \det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & -5 & 0 \\ 0 & 2-\lambda & 0 \\ -1 & 4 & 1-\lambda \end{bmatrix}\right) = (2-\lambda)^2(1-\lambda).$$

The eigenvalues are 1 and 2, where 2 has multiplicity 2. We leave it to the reader to find that $\left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right]$ is an eigenvector for the eigenvalue $\lambda = 1$.

Let us focus on $\lambda = 2$. We compute eigenvectors:

$$\vec{0} = (A - 2I)\vec{v} = \begin{bmatrix} 0 & -5 & 0 \\ 0 & 0 & 0 \\ -1 & 4 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

The first equation says that $v_2 = 0$, so the last equation is $-v_1 - v_3 = 0$. Let v_3 be the free variable to find that $v_1 = -v_3$. Perhaps let $v_3 = -1$ to find an eigenvector $\left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right]$. Problem is that setting v_3 to anything else just gets multiples of this vector and so we have a defect of 1. Let $\vec{v}_1$ be the eigenvector and let us look for a generalized eigenvector $\vec{v}_2$:

$$(A - 2I)\vec{v}_2 = \vec{v}_1,$$

or

$$\begin{bmatrix} 0 & -5 & 0 \\ 0 & 0 & 0 \\ -1 & 4 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

where we used a, b, c as components of $\vec{v}_2$ for simplicity. The first equation says $-5b = 1$ so $b = {-1}/{5}$. The second equation says nothing. The last equation is $-a + 4b - c = -1$, or

$a + \nicefrac{4}{5} + c = 1$, or $a + c = \nicefrac{1}{5}$. We let c be the free variable and we choose $c = 0$. We find $\vec{v}_2 = \left[\begin{smallmatrix} 1/5 \\ -1/5 \\ 0 \end{smallmatrix}\right]$.

The general solution is therefore,

$$\vec{x} = c_1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} e^{t} + c_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} e^{2t} + c_3 \left(\begin{bmatrix} 1/5 \\ -1/5 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} t \right) e^{2t}.$$

This machinery can also be generalized to higher multiplicities and higher defects. We will not go over this method in detail, but let us just sketch the ideas. Suppose that A has an eigenvalue λ of multiplicity m. We find vectors such that

$$(A - \lambda I)^k \vec{v} = \vec{0}, \qquad \text{but} \qquad (A - \lambda I)^{k-1} \vec{v} \neq \vec{0}.$$

Such vectors are called *generalized eigenvectors* (then $\vec{v}_1 = (A - \lambda I)^{k-1} \vec{v}$ is an eigenvector). For the eigenvector $\vec{v}_1$ there is a chain of generalized eigenvectors $\vec{v}_2$ through $\vec{v}_k$ such that:

$$\begin{aligned} (A - \lambda I)\vec{v}_1 &= \vec{0}, \\ (A - \lambda I)\vec{v}_2 &= \vec{v}_1, \\ &\vdots \\ (A - \lambda I)\vec{v}_k &= \vec{v}_{k-1}. \end{aligned}$$

Really once you find the $\vec{v}_k$ such that $(A - \lambda I)^k \vec{v}_k = \vec{0}$ but $(A - \lambda I)^{k-1} \vec{v}_k \neq \vec{0}$, you find the entire chain since you can compute the rest, $\vec{v}_{k-1} = (A - \lambda I)\vec{v}_k$, $\vec{v}_{k-2} = (A - \lambda I)\vec{v}_{k-1}$, etc. We form the linearly independent solutions

$$\begin{aligned} \vec{x}_1 &= \vec{v}_1 e^{\lambda t}, \\ \vec{x}_2 &= (\vec{v}_2 + \vec{v}_1 t)\, e^{\lambda t}, \\ &\vdots \\ \vec{x}_k &= \left(\vec{v}_k + \vec{v}_{k-1} t + \vec{v}_{k-2} \frac{t^2}{2} + \cdots + \vec{v}_2 \frac{t^{k-2}}{(k-2)!} + \vec{v}_1 \frac{t^{k-1}}{(k-1)!} \right) e^{\lambda t}. \end{aligned}$$

Recall that $k! = 1 \cdot 2 \cdot 3 \cdots (k-1) \cdot k$ is the factorial. If you have an eigenvalue of geometric multiplicity ℓ, you will have to find ℓ such chains (some of them might be short: just the single eigenvector equation). We go until we form m linearly independent solutions where m is the algebraic multiplicity. We don't quite know which specific eigenvectors go with which chain, so start by finding $\vec{v}_k$ first for the longest possible chain and go from there.

For example, if λ is an eigenvalue of A of algebraic multiplicity 3 and defect 2, then solve

$$(A - \lambda I)\vec{v}_1 = \vec{0}, \qquad (A - \lambda I)\vec{v}_2 = \vec{v}_1, \qquad (A - \lambda I)\vec{v}_3 = \vec{v}_2.$$

That is, find $\vec{v}_3$ such that $(A - \lambda I)^3 \vec{v}_3 = \vec{0}$, but $(A - \lambda I)^2 \vec{v}_3 \neq \vec{0}$. Then you are done as $\vec{v}_2 = (A - \lambda I)\vec{v}_3$ and $\vec{v}_1 = (A - \lambda I)\vec{v}_2$. The 3 linearly independent solutions are

$$\vec{x}_1 = \vec{v}_1 e^{\lambda t}, \qquad \vec{x}_2 = (\vec{v}_2 + \vec{v}_1 t)\, e^{\lambda t}, \qquad \vec{x}_3 = \left(\vec{v}_3 + \vec{v}_2 t + \vec{v}_1 \frac{t^2}{2}\right) e^{\lambda t}.$$

If, on the other hand, A has an eigenvalue λ of algebraic multiplicity 3 and defect 1, then solve

$$(A - \lambda I)\vec{v}_1 = \vec{0}, \qquad (A - \lambda I)\vec{v}_2 = \vec{0}, \qquad (A - \lambda I)\vec{v}_3 = \vec{v}_2.$$

Here $\vec{v}_1$ and $\vec{v}_2$ are actual honest eigenvectors, and $\vec{v}_3$ is a generalized eigenvector. So there are two chains. To solve, first find a $\vec{v}_3$ such that $(A - \lambda I)^2 \vec{v}_3 = \vec{0}$, but $(A - \lambda I)\vec{v}_3 \neq \vec{0}$. Then $\vec{v}_2 = (A - \lambda I)\vec{v}_3$ is going to be an eigenvector. Then solve for an eigenvector $\vec{v}_1$ that is linearly independent from $\vec{v}_2$. You get 3 linearly independent solutions

$$\vec{x}_1 = \vec{v}_1 e^{\lambda t}, \qquad \vec{x}_2 = \vec{v}_2 e^{\lambda t}, \qquad \vec{x}_3 = (\vec{v}_3 + \vec{v}_2 t)\, e^{\lambda t}.$$

2.8.3 Exercises

***Exercise* 2.8.2:** *Let* $A = \left[\begin{smallmatrix} 5 & -3 \\ 3 & -1 \end{smallmatrix}\right]$. *Find the general solution of* $\vec{x}' = A\vec{x}$.

***Exercise* 2.8.3:** *Let* $A = \left[\begin{smallmatrix} 5 & -4 & 4 \\ 0 & 3 & 0 \\ -2 & 4 & -1 \end{smallmatrix}\right]$.

a) What are the eigenvalues?

b) What is/are the defect(s) of the eigenvalue(s)?

c) Find the general solution of $\vec{x}' = A\vec{x}$.

***Exercise* 2.8.4:** *Let* $A = \left[\begin{smallmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right]$.

a) What are the eigenvalues?

b) What is/are the defect(s) of the eigenvalue(s)?

c) Find the general solution of $\vec{x}' = A\vec{x}$ *in two different ways and verify you get the same answer.*

***Exercise* 2.8.5:** *Let* $A = \left[\begin{smallmatrix} 0 & 1 & 2 \\ -1 & -2 & -2 \\ -4 & 4 & 7 \end{smallmatrix}\right]$.

a) What are the eigenvalues?

b) What is/are the defect(s) of the eigenvalue(s)?

c) Find the general solution of $\vec{x}' = A\vec{x}$.

***Exercise* 2.8.6:** *Let* $A = \begin{bmatrix} 0 & 4 & -2 \\ -1 & -4 & 1 \\ 0 & 0 & -2 \end{bmatrix}$.

a) *What are the eigenvalues?*

b) *What is/are the defect(s) of the eigenvalue(s)?*

c) *Find the general solution of* $\vec{x}' = A\vec{x}$.

***Exercise* 2.8.7:** *Let* $A = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 0 & 2 \\ -1 & -2 & 4 \end{bmatrix}$.

a) *What are the eigenvalues?*

b) *What is/are the defect(s) of the eigenvalue(s)?*

c) *Find the general solution of* $\vec{x}' = A\vec{x}$.

***Exercise* 2.8.8:** *Suppose that* A *is a* 2×2 *matrix with a repeated eigenvalue* λ. *Suppose that there are two linearly independent eigenvectors. Show that* $A = \lambda I$.

***Exercise* 2.8.101:** *Let* $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

a) *What are the eigenvalues?*

b) *What is/are the defect(s) of the eigenvalue(s)?*

c) *Find the general solution of* $\vec{x}\,' = A\vec{x}$.

***Exercise* 2.8.102:** *Let* $A = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix}$.

a) *What are the eigenvalues?*

b) *What is/are the defect(s) of the eigenvalue(s)?*

c) *Find the general solution of* $\vec{x}\,' = A\vec{x}$.

***Exercise* 2.8.103:** *Let* $A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & -1 & 9 \\ 0 & -1 & 5 \end{bmatrix}$.

a) *What are the eigenvalues?*

b) *What is/are the defect(s) of the eigenvalue(s)?*

c) *Find the general solution of* $\vec{x}\,' = A\vec{x}$.

***Exercise* 2.8.104:** *Let* $A = \left[\begin{smallmatrix} a & a \\ b & c \end{smallmatrix}\right]$, *where* a, b, *and* c *are unknowns. Suppose that* 5 *is a doubled eigenvalue of defect* 1, *and suppose that* $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ *is a corresponding eigenvector. Find* A *and show that there is only one such matrix* A.

Chapter 3

Nonlinear systems

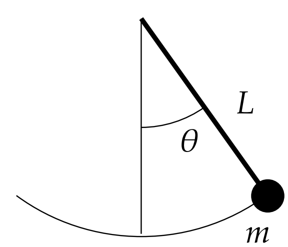
3.1 Linearization, critical points, and equilibria

Note: 1 lecture, §6.1–§6.2 in [EP], §9.2–§9.3 in [BD]

Except for a few brief detours in chapter 1, we considered mostly linear equations. Linear equations suffice in many applications, but in reality most phenomena require nonlinear equations. Nonlinear equations, however, are notoriously more difficult to understand than linear ones, and many strange new phenomena appear when we allow our equations to be nonlinear.

Not to worry, we did not waste all this time studying linear equations. Nonlinear equations can often be approximated by linear ones if we only need a solution "locally," for example, only for a short period of time, or only for certain parameters. Understanding linear equations can also give us qualitative understanding about a more general nonlinear problem. The idea is similar to what you did in calculus in trying to approximate a function by a line with the right slope.

In § **??** we looked at the pendulum of length L. The goal was to solve for the angle $\theta(t)$ as a function of the time t. The equation for the setup is the nonlinear equation



$$\theta'' + \frac{g}{L}\sin\theta = 0.$$

Instead of solving this equation, we solved the rather easier linear equation

$$\theta'' + \frac{g}{L}\theta = 0.$$

While the solution to the linear equation is not exactly what we were looking for, it is rather close to the original, as long as the angle θ is small and the time period involved is short.

You might ask: Why don't we just solve the nonlinear problem? Well, it might be very difficult, impractical, or impossible to solve analytically, depending on the equation in question. We may not even be interested in the actual solution, we might only be interested in some qualitative idea of what the solution is doing. For example, what happens as time goes to infinity?

3.1.1 Autonomous systems and phase plane analysis

We restrict our attention to a two-dimensional autonomous system

$$x' = f(x, y), \qquad y' = g(x, y),$$

where $f(x, y)$ and $g(x, y)$ are functions of two variables, and the derivatives are taken with respect to time t. Solutions are functions $x(t)$ and $y(t)$ such that

$$x'(t) = f\big(x(t), y(t)\big), \qquad y'(t) = g\big(x(t), y(t)\big).$$

The way we will analyze the system is very similar to § 1.6, where we studied a single autonomous equation. The ideas in two dimensions are the same, but the behavior can be far more complicated.

It may be best to think of the system of equations as the single vector equation

$$\begin{bmatrix} x \\ y \end{bmatrix}' = \begin{bmatrix} f(x, y) \\ g(x, y) \end{bmatrix}. \tag{3.1}$$

As in § 2.1 we draw the *phase portrait* (or *phase diagram*), where each point (x, y) corresponds to a specific state of the system. We draw the *vector field* given at each point (x, y) by the vector $\left[\begin{smallmatrix} f(x,y) \\ g(x,y) \end{smallmatrix}\right]$. And as before if we find solutions, we draw the trajectories by plotting all points $\big(x(t), y(t)\big)$ for a certain range of t.

Example 3.1.1: Consider the second order equation $x'' = -x + x^2$. Write this equation as a first order nonlinear system

$$x' = y, \qquad y' = -x + x^2.$$

The phase portrait with some trajectories is drawn in Figure 3.1 on the facing page.

From the phase portrait it should be clear that even this simple system has fairly complicated behavior. Some trajectories keep oscillating around the origin, and some go off towards infinity. We will return to this example often, and analyze it completely in this (and the next) section.

If we zoom into the diagram near a point where $\left[\begin{smallmatrix} f(x,y) \\ g(x,y) \end{smallmatrix}\right]$ is not zero, then nearby the arrows point generally in essentially that same direction and have essentially the same magnitude. In other words the behavior is not that interesting near such a point. We are of course assuming that $f(x, y)$ and $g(x, y)$ are continuous.

Let us concentrate on those points in the phase diagram above where the trajectories seem to start, end, or go around. We see two such points: $(0, 0)$ and $(1, 0)$. The trajectories seem to go around the point $(0, 0)$, and they seem to either go in or out of the point $(1, 0)$.

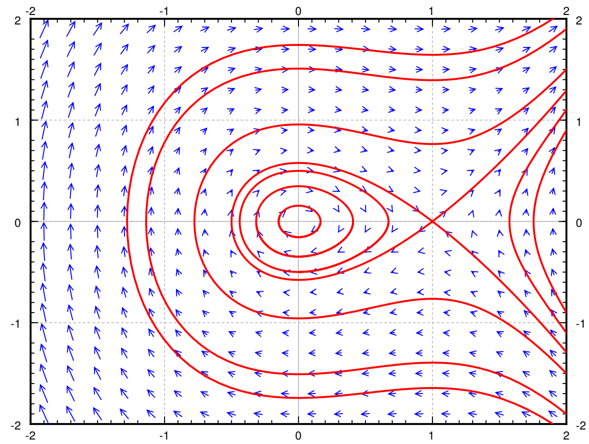


Figure 3.1: Phase portrait with some trajectories of $x' = y$, $y' = -x + x^2$.

These points are precisely those points where the derivatives of both x and y are zero. Let us define the *critical points* as the points (x, y) such that

$$\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix} = \vec{0}.$$

In other words, these are the points where both $f(x, y) = 0$ and $g(x, y) = 0$.

The critical points are where the behavior of the system is in some sense the most complicated. If $\left[\begin{smallmatrix} f(x,y) \\ g(x,y) \end{smallmatrix}\right]$ is zero, then nearby, the vector can point in any direction whatsoever. Also, the trajectories are either going towards, away from, or around these points, so if we are looking for long-term qualitative behavior of the system, we should look at what is happening near the critical points.

Critical points are also sometimes called *equilibria*, since we have so-called *equilibrium solutions* at critical points. If (x_0, y_0) is a critical point, then we have the solutions

$$x(t) = x_0, \quad y(t) = y_0.$$

In Example 3.1.1 on the preceding page, there are two equilibrium solutions:

$$x(t) = 0, \quad y(t) = 0, \qquad \text{and} \qquad x(t) = 1, \quad y(t) = 0.$$

Compare this discussion on equilibria to the discussion in § 1.6. The underlying concept is exactly the same.

3.1.2 Linearization

In § 2.5 we studied the behavior of a homogeneous linear system of two equations near a critical point. For a linear system of two variables given by an invertible matrix, the only critical point is the origin $(0, 0)$. Let us put the understanding we gained in that section to good use understanding what happens near critical points of nonlinear systems.

In calculus we learned to estimate a function by taking its derivative and linearizing. We work similarly with nonlinear systems of ODE. Suppose (x_0, y_0) is a critical point. First change variables to (u, v), so that $(u, v) = (0, 0)$ corresponds to (x_0, y_0). That is,

$$u = x - x_0, \qquad v = y - y_0.$$

Next we need to find the derivative. In multivariable calculus you may have seen that the several variables version of the derivative is the *Jacobian matrix**. The Jacobian matrix of the vector-valued function $\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}$ at (x_0, y_0) is

$$\begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \\ \frac{\partial g}{\partial x}(x_0, y_0) & \frac{\partial g}{\partial y}(x_0, y_0) \end{bmatrix}.$$

This matrix gives the best linear approximation as u and v (and therefore x and y) vary. We define the *linearization* of the equation (3.1) as the linear system

$$\begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \\ \frac{\partial g}{\partial x}(x_0, y_0) & \frac{\partial g}{\partial y}(x_0, y_0) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.$$

Example 3.1.2: Let us keep with the same equations as Example 3.1.1: $x' = y$, $y' = -x + x^2$. There are two critical points, $(0, 0)$ and $(1, 0)$. The Jacobian matrix at any point is

$$\begin{bmatrix} \frac{\partial f}{\partial x}(x, y) & \frac{\partial f}{\partial y}(x, y) \\ \frac{\partial g}{\partial x}(x, y) & \frac{\partial g}{\partial y}(x, y) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 + 2x & 0 \end{bmatrix}.$$

Therefore at $(0, 0)$, we have $u = x$ and $v = y$, and the linearization is

$$\begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.$$

At the point $(1, 0)$, we have $u = x - 1$ and $v = y$, and the linearization is

$$\begin{bmatrix} u \\ v \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.$$

The phase diagrams of the two linearizations at the point $(0, 0)$ and $(1, 0)$ are given in Figure 3.2 on the facing page. Note that the variables are now u and v. Compare Figure 3.2 with Figure 3.1 on the previous page, and look especially at the behavior near the critical points.

*Named for the German mathematician Carl Gustav Jacob Jacobi (1804–1851).

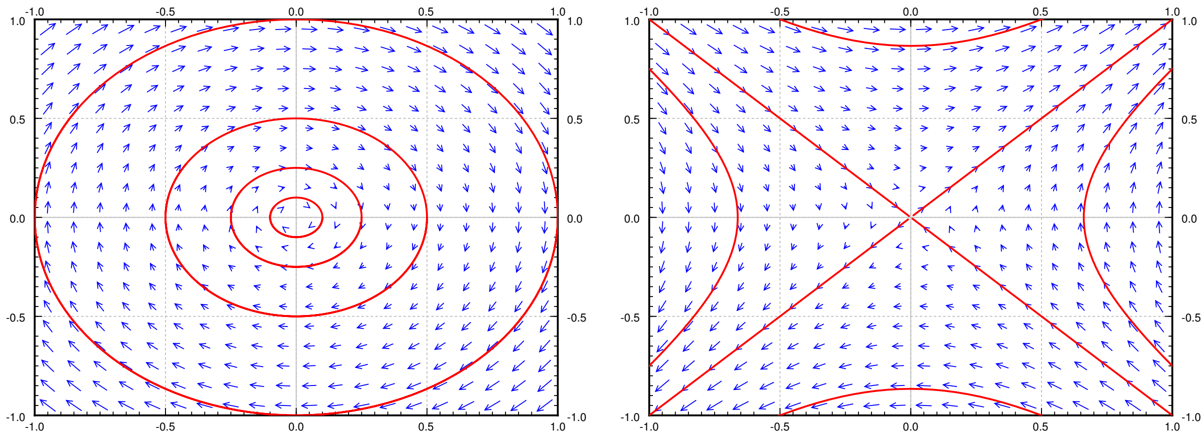


Figure 3.2: Phase diagram with some trajectories of linearizations at the critical points $(0,0)$ *(left) and* $(1,0)$ *(right) of* $x' = y$, $y' = -x + x^2$.

3.1.3 Exercises

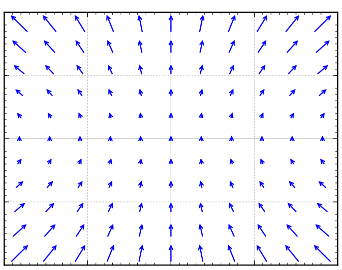
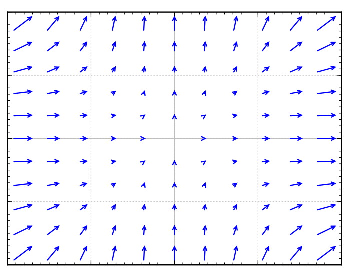
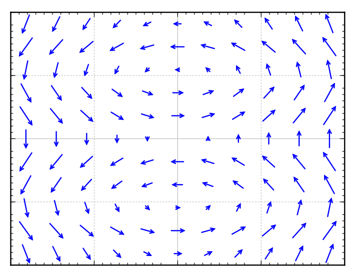
***Exercise* 3.1.1:** *Sketch the phase plane vector field for:*

a) $x' = x^2, \quad y' = y^2,$ *b)* $x' = (x-y)^2, \quad y' = -x,$ *c)* $x' = e^y, \quad y' = e^x.$

***Exercise* 3.1.2:** *Match systems*

1) $x' = x^2, \quad y' = y^2,$ *2)* $x' = xy, \quad y' = 1 + y^2,$ *3)* $x' = \sin(\pi y), \quad y' = x,$

to the vector fields below. Justify.

a)  *b)*  *c)* 

***Exercise* 3.1.3:** *Find the critical points and linearizations of the following systems.*

a) $x' = x^2 - y^2, \quad y' = x^2 + y^2 - 1,$ *b)* $x' = -y, \quad y' = 3x + yx^2,$

c) $x' = x^2 + y, \quad y' = y^2 + x.$

***Exercise* 3.1.4:** *For the following systems, verify they have critical point at* $(0,0)$*, and find the linearization at* $(0,0)$.

a) $x' = x + 2y + x^2 - y^2, \quad y' = 2y - x^2$ *b)* $x' = -y, \quad y' = x - y^3$

c) $x' = ax + by + f(x,y)$, $y' = cx + dy + g(x,y)$, *where* $f(0,0) = 0$, $g(0,0) = 0$, *and all first partial derivatives of* f *and* g *are also zero at* $(0,0)$, *that is,* $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = \frac{\partial g}{\partial x}(0,0) = \frac{\partial g}{\partial y}(0,0) = 0$.

***Exercise* 3.1.5:** *Take* $x' = (x-y)^2$, $y' = (x+y)^2$.

a) Find the set of critical points.

b) Sketch a phase diagram and describe the behavior near the critical point(s).

c) Find the linearization. Is it helpful in understanding the system?

***Exercise* 3.1.6:** *Take* $x' = x^2$, $y' = x^3$.

a) Find the set of critical points.

b) Sketch a phase diagram and describe the behavior near the critical point(s).

c) Find the linearization. Is it helpful in understanding the system?

***Exercise* 3.1.101:** *Find the critical points and linearizations of the following systems.*

a) $x' = \sin(\pi y) + (x-1)^2$, $y' = y^2 - y$, *b)* $x' = x + y + y^2$, $y' = x$,

c) $x' = (x-1)^2 + y$, $y' = x^2 + y$.

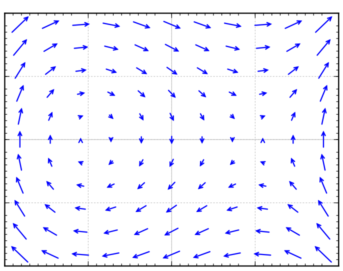
***Exercise* 3.1.102:** *Match systems*

1) $x' = y^2$, $y' = -x^2$, *2)* $x' = y$, $y' = (x-1)(x+1)$,

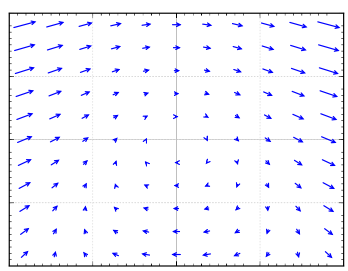
3) $x' = y + x^2$, $y' = -x$,

to the vector fields below. Justify.

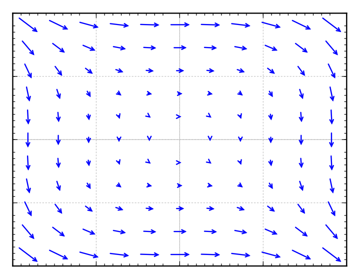
a)



b)



c)



***Exercise* 3.1.103:** *The idea of critical points and linearization works in higher dimensions as well. You simply make the Jacobian matrix bigger by adding more functions and more variables. For the following system of 3 equations find the critical points and their linearizations:*

$$x' = x + z^2, \qquad y' = z^2 - y, \qquad z' = z + x^2.$$

***Exercise* 3.1.104:** *Any two-dimensional non-autonomous system* $x' = f(x,y,t)$, $y' = g(x,y,t)$ *can be written as a three-dimensional autonomous system (three equations). Write down this autonomous system using the variables* u, v, w.

3.2 Stability and classification of isolated critical points

Note: 1.5–2 lectures, §6.1–§6.2 in [EP], §9.2–§9.3 in [BD]

3.2.1 Isolated critical points and almost linear systems

A critical point is *isolated* if it is the only critical point in some small "neighborhood" of the point. That is, if we zoom in far enough it is the only critical point we see. In the example above, the critical point was isolated. If on the other hand there would be a whole curve of critical points, then it would not be isolated.

A system is called *almost linear* at a critical point (x_0, y_0), if the critical point is isolated and the Jacobian matrix at the point is invertible, or equivalently if the linearized system has an isolated critical point. In such a case, the nonlinear terms are very small and the system behaves like its linearization, at least if we are close to the critical point.

For example, the system in Examples 3.1.1 and 3.1.2 has two isolated critical points $(0, 0)$ and $(0, 1)$, and is almost linear at both critical points as the Jacobian matrices at both points, $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, are invertible.

On the other hand, the system $x' = x^2$, $y' = y^2$ has an isolated critical point at $(0, 0)$, however the Jacobian matrix

$$\begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix}$$

is zero when $(x, y) = (0, 0)$. So the system is not almost linear. Even a worse example is the system $x' = x$, $y' = x^2$, which does not have isolated critical points; x' and y' are both zero whenever $x = 0$, that is, the entire y-axis.

Fortunately, most often critical points are isolated, and the system is almost linear at the critical points. So if we learn what happens there, we will have figured out the majority of situations that arise in applications.

3.2.2 Stability and classification of isolated critical points

Once we have an isolated critical point, the system is almost linear at that critical point, and we computed the associated linearized system, we can classify what happens to the solutions. We more or less use the classification for linear two-variable systems from § 2.5, with one minor caveat. Let us list the behaviors depending on the eigenvalues of the Jacobian matrix at the critical point in Table 3.1 on the following page. This table is very similar to Table 2.1 on page 102, with the exception of missing "center" points. We will discuss centers later, as they are more complicated.

In the third column, we mark points as *asymptotically stable* or *unstable*. Formally, a *stable critical point* (x_0, y_0) is one where given any small distance ϵ to (x_0, y_0), and any initial condition within a perhaps smaller radius around (x_0, y_0), the trajectory of the system never goes further away from (x_0, y_0) than ϵ. An *unstable critical point* is one that is not

Eigenvalues of the Jacobian matrix	Behavior	Stability
real and both positive	source / unstable node	unstable
real and both negative	sink / stable node	asymptotically stable
real and opposite signs	saddle	unstable
complex with positive real part	spiral source	unstable
complex with negative real part	spiral sink	asymptotically stable

Table 3.1: Behavior of an almost linear system near an isolated critical point.

stable. Informally, a point is stable if we start close to a critical point and follow a trajectory we either go towards, or at least not away from, this critical point.

A stable critical point (x_0, y_0) is called *asymptotically stable* if given any initial condition sufficiently close to (x_0, y_0) and any solution $\big(x(t), y(t)\big)$ satisfying that condition, then

$$\lim_{t\to\infty} \big(x(t), y(t)\big) = (x_0, y_0).$$

That is, the critical point is asymptotically stable if any trajectory for a sufficiently close initial condition goes towards the critical point (x_0, y_0).

Example 3.2.1: Consider $x' = -y - x^2$, $y' = -x + y^2$. See Figure 3.3 on the facing page for the phase diagram. Let us find the critical points. These are the points where $-y - x^2 = 0$ and $-x + y^2 = 0$. The first equation means $y = -x^2$, and so $y^2 = x^4$. Plugging into the second equation we obtain $-x + x^4 = 0$. Factoring we obtain $x(1 - x^3) = 0$. Since we are looking only for real solutions we get either $x = 0$ or $x = 1$. Solving for the corresponding y using $y = -x^2$, we get two critical points, one being $(0, 0)$ and the other being $(1, -1)$. Clearly the critical points are isolated.

Let us compute the Jacobian matrix:

$$\begin{bmatrix} -2x & -1 \\ -1 & 2y \end{bmatrix}.$$

At the point $(0, 0)$ we get the matrix $\left[\begin{smallmatrix} 0 & -1 \\ -1 & 0 \end{smallmatrix}\right]$ and so the two eigenvalues are 1 and -1. As the matrix is invertible, the system is almost linear at $(0, 0)$. As the eigenvalues are real and of opposite signs, we get a saddle point, which is an unstable equilibrium point.

At the point $(1, -1)$ we get the matrix $\left[\begin{smallmatrix} -2 & -1 \\ -1 & -2 \end{smallmatrix}\right]$ and computing the eigenvalues we get -1, -3. The matrix is invertible, and so the system is almost linear at $(1, -1)$. As we have real eigenvalues and both negative, the critical point is a sink, and therefore an asymptotically stable equilibrium point. That is, if we start with any point (x_i, y_i) close to $(1, -1)$ as an initial condition and plot a trajectory, it approaches $(1, -1)$. In other words,

$$\lim_{t\to\infty} \big(x(t), y(t)\big) = (1, -1).$$

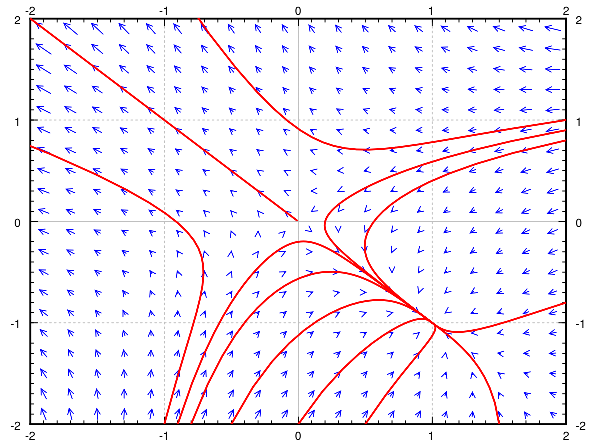


Figure 3.3: The phase portrait with few sample trajectories of $x' = -y - x^2$, $y' = -x + y^2$.

As you can see from the diagram, this behavior is true even for some initial points quite far from $(1,-1)$, but it is definitely not true for all initial points.

Example 3.2.2: Let us look at $x' = y + y^2 e^x$, $y' = x$. First let us find the critical points. These are the points where $y + y^2 e^x = 0$ and $x = 0$. Simplifying we get $0 = y + y^2 = y(y+1)$. So the critical points are $(0,0)$ and $(0,-1)$, and hence are isolated. Let us compute the Jacobian matrix:

$$\begin{bmatrix} y^2 e^x & 1 + 2ye^x \\ 1 & 0 \end{bmatrix}.$$

At the point $(0,0)$ we get the matrix $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ and so the two eigenvalues are 1 and -1. As the matrix is invertible, the system is almost linear at $(0,0)$. And, as the eigenvalues are real and of opposite signs, we get a saddle point, which is an unstable equilibrium point.

At the point $(0,-1)$ we get the matrix $\left[\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ whose eigenvalues are $\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$. The matrix is invertible, and so the system is almost linear at $(0,-1)$. As we have complex eigenvalues with positive real part, the critical point is a spiral source, and therefore an unstable equilibrium point.

See Figure 3.4 on the next page for the phase diagram. Notice the two critical points, and the behavior of the arrows in the vector field around these points.

3.2.3 The trouble with centers

Recall, a linear system with a center means that trajectories travel in closed elliptical orbits in some direction around the critical point. Such a critical point we call a *center* or a *stable center*. It is not an asymptotically stable critical point, as the trajectories never approach the critical point, but at least if you start sufficiently close to the critical point, you stay close to the critical point. The simplest example of such behavior is the linear system with a center. Another example is the critical point $(0,0)$ in Example 3.1.1 on page 126.

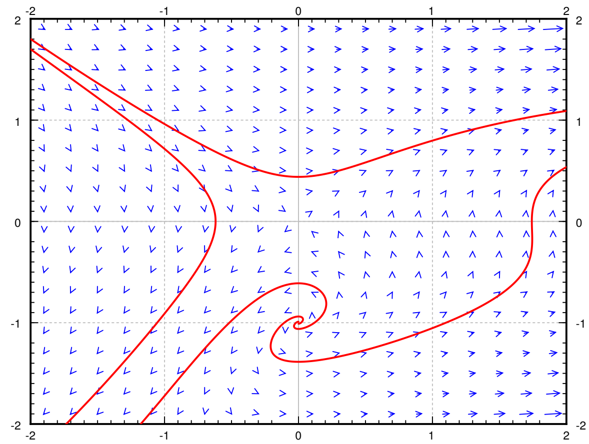


Figure 3.4: The phase portrait with few sample trajectories of $x' = y + y^2 e^x$, $y' = x$.

The trouble with a center in a nonlinear system is that whether the trajectory goes towards or away from the critical point is governed by the sign of the real part of the eigenvalues of the Jacobian matrix, and the Jacobian matrix in a nonlinear system changes from point to point. Since this real part is zero at the critical point itself, it can have either sign nearby, meaning the trajectory could be pulled towards or away from the critical point.

Example 3.2.3: An example of such a problematic behavior is the system $x' = y$, $y' = -x + y^3$. The only critical point is the origin $(0,0)$. The Jacobian matrix is

$$\begin{bmatrix} 0 & 1 \\ -1 & 3y^2 \end{bmatrix}.$$

At $(0,0)$ the Jacobian matrix is $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, which has eigenvalues $\pm i$. So the linearization has a center.

Using the quadratic equation, the eigenvalues of the Jacobian matrix at any point (x, y) are

$$\lambda = \frac{3}{2} y^2 \pm i \frac{\sqrt{4 - 9y^4}}{2}.$$

At any point where $y \neq 0$ (so at most points near the origin), the eigenvalues have a positive real part (y^2 can never be negative). This positive real part pulls the trajectory away from the origin. A sample trajectory for an initial condition near the origin is given in Figure 3.5 on the next page.

The moral of the example is that further analysis is needed when the linearization has a center. The analysis will in general be more complicated than in the example above, and is more likely to involve case-by-case consideration. Such a complication should not be surprising to you. By now in your mathematical career, you have seen many places where a simple test is inconclusive, recall for example the second derivative test for maxima or minima, and requires more careful, and perhaps ad hoc analysis of the situation.

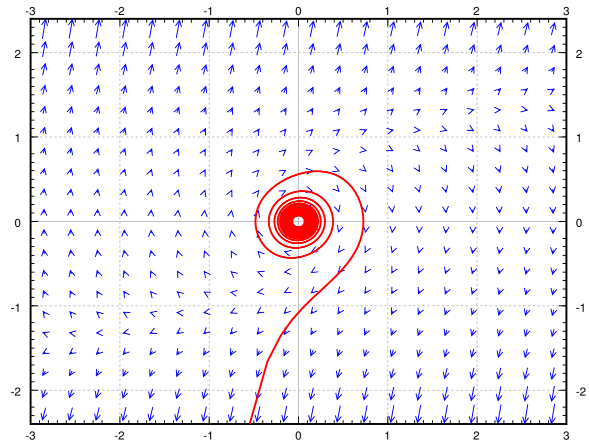


Figure 3.5: An unstable critical point (spiral source) at the origin for $x' = y$, $y' = -x + y^3$, even if the linearization has a center.

3.2.4 Conservative equations

An equation of the form

$$x'' + f(x) = 0,$$

where $f(x)$ is an arbitrary function, is called a *conservative equation*. For example, the pendulum equation is a conservative equation. The equations are conservative as there is no friction in the system so the energy in the system is "conserved." Let us write this equation as a system of nonlinear ODE.

$$x' = y, \qquad y' = -f(x).$$

These types of equations have the advantage that we can solve for their trajectories easily.

The trick is to first think of y as a function of x for a moment. Then use the chain rule

$$x'' = y' = \frac{dy}{dx}x' = y\frac{dy}{dx},$$

where the prime indicates a derivative with respect to t. We obtain $y\frac{dy}{dx} + f(x) = 0$. We integrate with respect to x to get $\int y\frac{dy}{dx}\,dx + \int f(x)\,dx = C$. In other words

$$\frac{1}{2}y^2 + \int f(x)\,dx = C.$$

We obtained an implicit equation for the trajectories, with different C giving different trajectories. The value of C is conserved on any trajectory. This expression is sometimes called the *Hamiltonian* or the energy of the system. If you look back to § 1.8, you will notice that $y\frac{dy}{dx} + f(x) = 0$ is an exact equation, and we just found a potential function.

Example 3.2.4: Let us find the trajectories for the equation $x'' + x - x^2 = 0$, which is the equation from Example 3.1.1 on page 126. The corresponding first order system is

$$x' = y, \qquad y' = -x + x^2.$$

Trajectories satisfy

$$\frac{1}{2}y^2 + \frac{1}{2}x^2 - \frac{1}{3}x^3 = C.$$

We solve for y

$$y = \pm\sqrt{-x^2 + \frac{2}{3}x^3 + 2C}.$$

Plotting these graphs we get exactly the trajectories in Figure 3.1 on page 127. In particular we notice that near the origin the trajectories are *closed curves*: they keep going around the origin, never spiraling in or out. Therefore we discovered a way to verify that the critical point at $(0, 0)$ is a stable center. The critical point at $(0, 1)$ is a saddle as we already noticed. This example is typical for conservative equations.

Consider an arbitrary conservative equation $x'' + f(x) = 0$. All critical points occur when $y = 0$ (the x-axis), that is when $x' = 0$. The critical points are those points on the x-axis where $f(x) = 0$. The trajectories are given by

$$y = \pm\sqrt{-2\int f(x)\,dx + 2C}.$$

So all trajectories are mirrored across the x-axis. In particular, there can be no spiral sources nor sinks. The Jacobian matrix is

$$\begin{bmatrix} 0 & 1 \\ -f'(x) & 0 \end{bmatrix}.$$

The critical point is almost linear if $f'(x) \neq 0$ at the critical point. Let J denote the Jacobian matrix. The eigenvalues of J are solutions to

$$0 = \det(J - \lambda I) = \lambda^2 + f'(x).$$

Therefore $\lambda = \pm\sqrt{-f'(x)}$. In other words, either we get real eigenvalues of opposite signs (if $f'(x) < 0$), or we get purely imaginary eigenvalues (if $f'(x) > 0$). There are only two possibilities for critical points, either an *unstable saddle point*, or a *stable center*. There are never any sinks or sources.

3.2.5 Exercises

Exercise* 3.2.1*: *For the systems below, find and classify the critical points, also indicate if the equilibria are stable, asymptotically stable, or unstable.*

a) $x' = -x + 3x^2, y' = -y$

b) $x' = x^2 + y^2 - 1, y' = x$

c) $x' = ye^x, y' = y - x + y^2$

***Exercise* 3.2.2:** *Find the implicit equations of the trajectories of the following conservative systems. Next find their critical points (if any) and classify them.*

a) $x'' + x + x^3 = 0$ *b)* $\theta'' + \sin\theta = 0$

c) $z'' + (z-1)(z+1) = 0$ *d)* $x'' + x^2 + 1 = 0$

***Exercise* 3.2.3:** *Find and classify the critical point(s) of* $x' = -x^2$, $y' = -y^2$.

***Exercise* 3.2.4:** *Suppose* $x' = -xy$, $y' = x^2 - 1 - y$.

a) Show there are two spiral sinks at $(-1, 0)$ *and* $(1, 0)$.

b) For any initial point of the form $(0, y_0)$, *find what is the trajectory.*

c) Can a trajectory starting at (x_0, y_0) *where* $x_0 > 0$ *spiral into the critical point at* $(-1, 0)$? *Why or why not?*

***Exercise* 3.2.5:** *In the example* $x' = y$, $y' = y^3 - x$ *show that for any trajectory, the distance from the origin is an increasing function. Conclude that the origin behaves like is a spiral source. Hint: Consider* $f(t) = \big(x(t)\big)^2 + \big(y(t)\big)^2$ *and show it has positive derivative.*

***Exercise* 3.2.6:** *Suppose* f *is always positive. Find the trajectories of* $x'' + f(x') = 0$. *Are there any critical points?*

***Exercise* 3.2.7:** *Suppose that* $x' = f(x, y)$, $y' = g(x, y)$. *Suppose that* $g(x, y) > 1$ *for all* x *and* y. *Are there any critical points? What can we say about the trajectories at* t *goes to infinity?*

***Exercise* 3.2.101:** *For the systems below, find and classify the critical points.*

a) $x' = -x + x^2$, $y' = y$ *b)* $x' = y - y^2 - x$, $y' = -x$ *c)* $x' = xy$, $y' = x + y - 1$

***Exercise* 3.2.102:** *Find the implicit equations of the trajectories of the following conservative systems. Next find their critical points (if any) and classify them.*

a) $x'' + x^2 = 4$ *b)* $x'' + e^x = 0$ *c)* $x'' + (x+1)e^x = 0$

***Exercise* 3.2.103:** *The conservative system* $x'' + x^3 = 0$ *is not almost linear. Classify its critical point(s) nonetheless.*

***Exercise* 3.2.104:** *Derive an analogous classification of critical points for equations in one dimension, such as* $x' = f(x)$ *based on the derivative. A point* x_0 *is critical when* $f(x_0) = 0$ *and almost linear if in addition* $f'(x_0) \neq 0$. *Figure out if the critical point is stable or unstable depending on the sign of* $f'(x_0)$. *Explain. Hint: see* § 1.6.

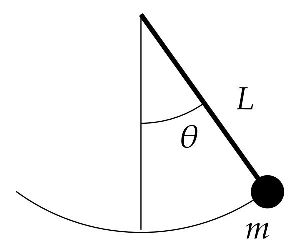
3.3 Applications of nonlinear systems

Note: 2 lectures, §6.3–§6.4 in [EP], §9.3, §9.5 in [BD]

In this section we study two very standard examples of nonlinear systems. First, we look at the nonlinear pendulum equation. We saw the pendulum equation's linearization before, but we noted it was only valid for small angles and short times. Now we find out what happens for large angles. Next, we look at the predator-prey equation, which finds various applications in modeling problems in biology, chemistry, economics, and elsewhere.

3.3.1 Pendulum

The first example we study is the pendulum equation $\theta'' + \frac{g}{L}\sin\theta = 0$. Here, θ is the angular displacement, g is the gravitational acceleration, and L is the length of the pendulum. In this equation we disregard friction, so we are talking about an idealized pendulum.



This equation is a conservative equation, so we can use our analysis of conservative equations from the previous section. Let us change the equation to a two-dimensional system in variables (θ, ω) by introducing the new variable ω:

$$\begin{bmatrix} \theta \\ \omega \end{bmatrix}' = \begin{bmatrix} \omega \\ -\frac{g}{L}\sin\theta \end{bmatrix}.$$

The critical points of this system are when $\omega = 0$ and $-\frac{g}{L}\sin\theta = 0$, or in other words if $\sin\theta = 0$. So the critical points are when $\omega = 0$ and θ is a multiple of π. That is, the points are $\ldots(-2\pi,0),(-\pi,0),(0,0),(\pi,0),(2\pi,0)\ldots$. While there are infinitely many critical points, they are all isolated. Let us compute the Jacobian matrix:

$$\begin{bmatrix} \frac{\partial}{\partial\theta}\Big(\omega\Big) & \frac{\partial}{\partial\omega}\Big(\omega\Big) \\ \frac{\partial}{\partial\theta}\Big(-\frac{g}{L}\sin\theta\Big) & \frac{\partial}{\partial\omega}\Big(-\frac{g}{L}\sin\theta\Big) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{L}\cos\theta & 0 \end{bmatrix}.$$

For conservative equations, there are two types of critical points. Either stable centers, or saddle points. The eigenvalues of the Jacobian matrix are $\lambda = \pm\sqrt{-\frac{g}{L}\cos\theta}$.

The eigenvalues are going to be real when $\cos\theta < 0$. This happens at the odd multiples of π. The eigenvalues are going to be purely imaginary when $\cos\theta > 0$. This happens at the even multiples of π. Therefore the system has a stable center at the points $\ldots(-2\pi,0),(0,0),(2\pi,0)\ldots$, and it has an unstable saddle at the points $\ldots(-3\pi,0),(-\pi,0),(\pi,0),(3\pi,0)\ldots$. Look at the phase diagram in Figure 3.6 on the facing page, where for simplicity we let $\frac{g}{L} = 1$.

In the linearized equation we have only a single critical point, the center at $(0,0)$. Now we see more clearly what we meant when we said the linearization is good for small

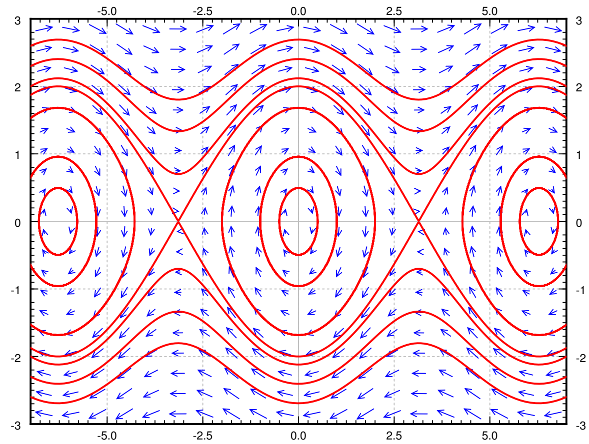


Figure 3.6: Phase plane diagram and some trajectories of the nonlinear pendulum equation.

angles. The horizontal axis is the deflection angle. The vertical axis is the angular velocity of the pendulum. Suppose we start at $\theta = 0$ (no deflection), and we start with a small angular velocity ω. Then the trajectory keeps going around the critical point $(0, 0)$ in an approximate circle. This corresponds to short swings of the pendulum back and forth. When θ stays small, the trajectories really look like circles and hence are very close to our linearization.

When we give the pendulum a big enough push, it goes across the top and keeps spinning about its axis. This behavior corresponds to the wavy curves that do not cross the horizontal axis in the phase diagram. Let us suppose we look at the top curves, when the angular velocity ω is large and positive. Then the pendulum is going around and around its axis. The velocity is going to be large when the pendulum is near the bottom, and the velocity is the smallest when the pendulum is close to the top of its loop.

At each critical point, there is an equilibrium solution. Consider the solution $\theta = 0$; the pendulum is not moving and is hanging straight down. This is a stable place for the pendulum to be, hence this is a *stable* equilibrium.

The other type of equilibrium solution is at the unstable point, for example $\theta = \pi$. Here the pendulum is upside down. Sure you can balance the pendulum this way and it will stay, but this is an *unstable* equilibrium. Even the tiniest push will make the pendulum start swinging wildly.

See Figure 3.7 on the next page for a diagram. The first picture is the stable equilibrium $\theta = 0$. The second picture corresponds to those "almost circles" in the phase diagram around $\theta = 0$ when the angular velocity is small. The next picture is the unstable equilibrium $\theta = \pi$. The last picture corresponds to the wavy lines for large angular velocities.

The quantity

$$\frac{1}{2}\omega^2 - \frac{g}{L}\cos\theta$$

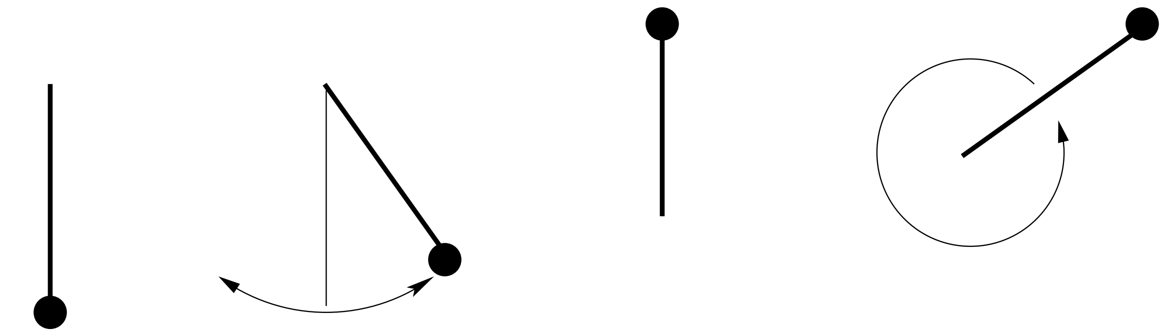


Figure 3.7: Various possibilities for the motion of the pendulum.

is conserved by any solution. This is the energy or the Hamiltonian of the system.

We have a conservative equation and so (exercise) the trajectories are given by

$$\omega = \pm\sqrt{\frac{2g}{L}\cos\theta + C},$$

for various values of C. Let us look at the initial condition of $(\theta_0, 0)$, that is, we take the pendulum to angle θ_0, and just let it go (initial angular velocity 0). We plug the initial conditions into the above and solve for C to obtain

$$C = -\frac{2g}{L}\cos\theta_0.$$

Thus the expression for the trajectory is

$$\omega = \pm\sqrt{\frac{2g}{L}}\sqrt{\cos\theta - \cos\theta_0}.$$

Let us figure out the period. That is, the time it takes for the pendulum to swing back and forth. We notice that the trajectory about the origin in the phase plane is symmetric about both the θ and the ω-axis. That is, in terms of θ, the time it takes from θ_0 to $-\theta_0$ is the same as it takes from $-\theta_0$ back to θ_0. Furthermore, the time it takes from $-\theta_0$ to 0 is the same as to go from 0 to θ_0. Therefore, let us find how long it takes for the pendulum to go from angle 0 to angle θ_0, which is a quarter of the full oscillation and then multiply by 4.

We figure out this time by finding $\frac{dt}{d\theta}$ and integrating from 0 to θ_0. The period is four times this integral. Let us stay in the region where ω is positive. Since $\omega = \frac{d\theta}{dt}$, inverting we get

$$\frac{dt}{d\theta} = \sqrt{\frac{L}{2g}}\frac{1}{\sqrt{\cos\theta - \cos\theta_0}}.$$

Therefore the period T is given by

$$T = 4\sqrt{\frac{L}{2g}} \int_0^{\theta_0} \frac{1}{\sqrt{\cos\theta - \cos\theta_0}}\, d\theta.$$

The integral is an improper integral, and we cannot in general evaluate it symbolically. We must resort to numerical approximation if we want to compute a particular T.

Recall from § **??**, the linearized equation $\theta'' + \frac{g}{L}\theta = 0$ has period

$$T_{\text{linear}} = 2\pi\sqrt{\frac{L}{g}}.$$

We plot T, T_{linear}, and the relative error $\frac{T - T_{\text{linear}}}{T}$ in Figure 3.8. The relative error says how far is our approximation from the real period percentage-wise. Note that T_{linear} is simply a constant, it does not change with the initial angle θ_0. The actual period T gets larger and larger as θ_0 gets larger. Notice how the relative error is small when θ_0 is small. It is still only 15% when $\theta_0 = \frac{\pi}{2}$, that is, a 90 degree angle. The error is 3.8% when starting at $\frac{\pi}{4}$, a 45 degree angle. At a 5 degree initial angle, the error is only 0.048%.

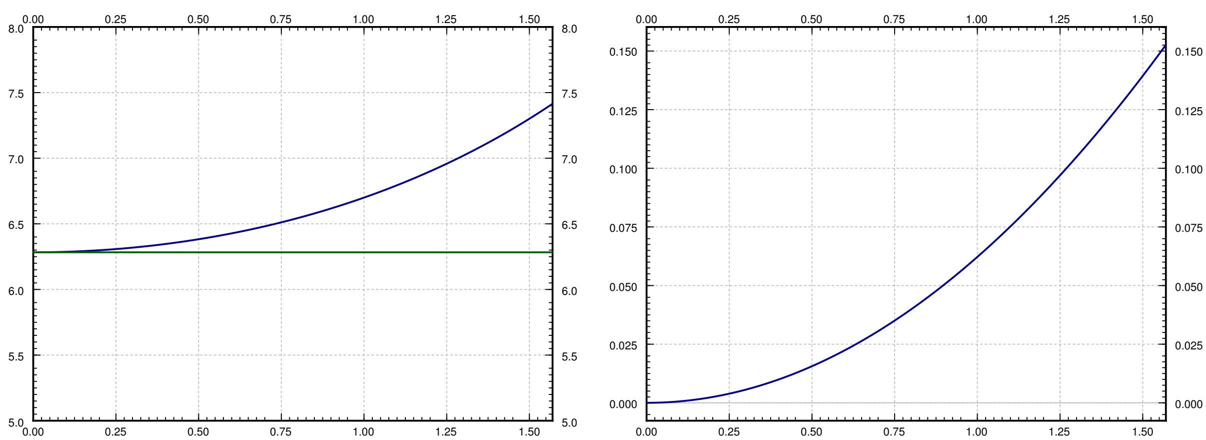


Figure 3.8: The plot of T and T_{linear} with $\frac{g}{L} = 1$ (left), and the plot of the relative error $\frac{T - T_{linear}}{T}$ (right), for θ_0 between 0 and $\pi/2$.

While it is not immediately obvious from the formula, it is true that

$$\lim_{\theta_0 \uparrow \pi} T = \infty.$$

That is, the period goes to infinity as the initial angle approaches the unstable equilibrium point. So if we put the pendulum almost upside down it may take a very long time before it gets down. This is consistent with the limiting behavior, where the exactly upside down pendulum never makes an oscillation, so we could think of that as infinite period.

3.3.2 Predator-prey or Lotka–Volterra systems

One of the most common simple applications of nonlinear systems are the so-called *predator-prey* or *Lotka–Volterra*[*] systems. For example, these systems arise when two species interact, one as the prey and one as the predator. It is then no surprise that the equations also see applications in economics. The system also arises in chemical reactions. In biology, this system of equations explains the natural periodic variations of populations of different species in nature. Before the application of differential equations, these periodic variations in the population baffled biologists.

We keep with the classical example of hares and foxes in a forest, it is the easiest to understand.

$$x = \text{\# of hares (the prey)},$$
$$y = \text{\# of foxes (the predator)}.$$

When there are a lot of hares, there is plenty of food for the foxes, so the fox population grows. However, when the fox population grows, the foxes eat more hares, so when there are lots of foxes, the hare population should go down, and vice versa. The Lotka–Volterra model proposes that this behavior is described by the system of equations

$$x' = (a - by)x,$$
$$y' = (cx - d)y,$$

where a, b, c, d are some parameters that describe the interaction of the foxes and hares[†]. In this model, these are all positive numbers.

Let us analyze the idea behind this model. The model is a slightly more complicated idea based on the exponential population model. First expand,

$$x' = (a - by)x = ax - byx.$$

The hares are expected to simply grow exponentially in the absence of foxes, that is where the ax term comes in, the growth in population is proportional to the population itself. We are assuming the hares always find enough food and have enough space to reproduce. However, there is another component $-byx$, that is, the population also is decreasing proportionally to the number of foxes. Together we can write the equation as $(a - by)x$, so it is like exponential growth or decay but the constant depends on the number of foxes.

The equation for foxes is very similar, expand again

$$y' = (cx - d)y = cxy - dy.$$

The foxes need food (hares) to reproduce: the more food, the bigger the rate of growth, hence the cxy term. On the other hand, there are natural deaths in the fox population, and hence the $-dy$ term.

---

[*]Named for the American mathematician, chemist, and statistician Alfred James Lotka (1880–1949) and the Italian mathematician and physicist Vito Volterra (1860–1940).

[†]This interaction does not end well for the hare.

Without further delay, let us start with an explicit example. Suppose the equations are

$$x' = (0.4 - 0.01y)x, \qquad y' = (0.003x - 0.3)y.$$

See Figure 3.9 for the phase portrait. In this example it makes sense to also plot x and y as graphs with respect to time. Therefore the second graph in Figure 3.9 is the graph of x and y on the vertical axis (the prey x is the thinner line with taller peaks), against time on the horizontal axis. The particular solution graphed was with initial conditions of 20 foxes and 50 hares.

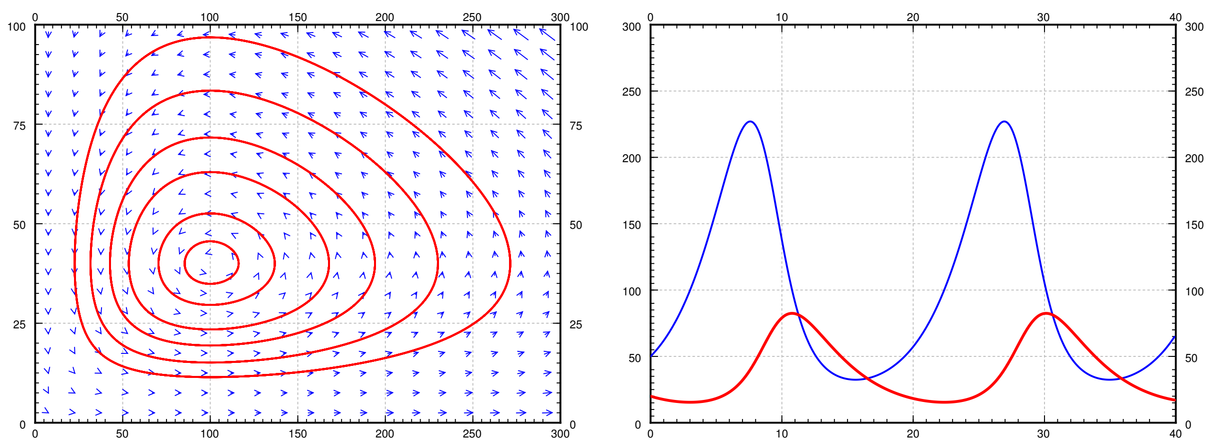


Figure 3.9: The phase portrait (left) and graphs of x and y for a sample solution (right).

Let us analyze what we see on the graphs. We work in the general setting rather than putting in specific numbers. We start with finding the critical points. Set $(a - by)x = 0$, and $(cx - d)y = 0$. The first equation is satisfied if either $x = 0$ or $y = a/b$. If $x = 0$, the second equation implies $y = 0$. If $y = a/b$, the second equation implies $x = d/c$. There are two equilibria: at $(0, 0)$ when there are no animals at all, and at $(d/c, a/b)$. In our specific example $x = d/c = 100$, and $y = a/b = 40$. This is the point where there are 100 hares and 40 foxes.

We compute the Jacobian matrix:

$$\begin{bmatrix} a - by & -bx \\ cy & cx - d \end{bmatrix}.$$

At the origin $(0, 0)$ we get the matrix $\left[\begin{smallmatrix} a & 0 \\ 0 & -d \end{smallmatrix}\right]$, so the eigenvalues are a and $-d$, hence real and of opposite signs. So the critical point at the origin is a saddle. This makes sense. If you started with some foxes but no hares, then the foxes would go extinct, that is, you would approach the origin. If you started with no foxes and a few hares, then the hares would keep multiplying without check, and so you would go away from the origin.

OK, how about the other critical point at $(d/c, a/b)$. Here the Jacobian matrix becomes

$$\begin{bmatrix} 0 & -\frac{bd}{c} \\ \frac{ac}{b} & 0 \end{bmatrix}.$$

The eigenvalues satisfy $\lambda^2 + ad = 0$. In other words, $\lambda = \pm i\sqrt{ad}$. The eigenvalues being purely imaginary, we are in the case where we cannot quite decide using only linearization. We could have a stable center, spiral sink, or a spiral source. That is, the equilibrium could be asymptotically stable, stable, or unstable. Of course I gave you a picture above that seems to imply it is a stable center. But never trust a picture only. Perhaps the oscillations are getting larger and larger, but only *very* slowly. Of course this would be bad as it would imply something will go wrong with our population sooner or later. And I only graphed a very specific example with very specific trajectories.

How can we be sure we are in the stable situation? As we said before, in the case of purely imaginary eigenvalues, we have to do a bit more work. Previously we found that for conservative systems, there was a certain quantity that was conserved on the trajectories, and hence the trajectories had to go in closed loops. We can use a similar technique here. We just have to figure out what is the conserved quantity. After some trial and error we find the constant

$$C = \frac{y^a x^d}{e^{cx+by}} = y^a x^d e^{-cx-by}$$

is conserved. Such a quantity is called the *constant of motion*. Let us check C really is a constant of motion. How do we check, you say? Well, a constant is something that does not change with time, so let us compute the derivative with respect to time:

$$C' = ay^{a-1}y'x^d e^{-cx-by} + y^a dx^{d-1}x' e^{-cx-by} + y^a x^d e^{-cx-by}(-cx' - by').$$

Our equations give us what x' and y' are so let us plug those in:

$$\begin{aligned} C' &= ay^{a-1}(cx-d)yx^d e^{-cx-by} + y^a dx^{d-1}(a-by)xe^{-cx-by} \\ &\qquad + y^a x^d e^{-cx-by}\big(-c(a-by)x - b(cx-d)y\big) \\ &= y^a x^d e^{-cx-by}\Big(a(cx-d) + d(a-by) + \big(-c(a-by)x - b(cx-d)y\big)\Big) \\ &= 0. \end{aligned}$$

So along the trajectories C is constant. In fact, the expression $C = \frac{y^a x^d}{e^{cx+by}}$ gives us an implicit equation for the trajectories. In any case, once we have found this constant of motion, it must be true that the trajectories are simple curves, that is, the level curves of $\frac{y^a x^d}{e^{cx+by}}$. It turns out, the critical point at $(d/c, a/b)$ is a maximum for C (left as an exercise). So $(d/c, a/b)$ is a stable equilibrium point, and we do not have to worry about the foxes and hares going extinct or their populations exploding.

One blemish on this wonderful model is that the number of foxes and hares are discrete quantities and we are modeling with continuous variables. Our model has no problem with there being 0.1 fox in the forest for example, while in reality that makes no sense. The approximation is a reasonable one as long as the number of foxes and hares are large, but it does not make much sense for small numbers. One must be careful in interpreting any results from such a model.

An interesting consequence (perhaps counterintuitive) of this model is that adding animals to the forest might lead to extinction, because the variations will get too big, and one of the populations will get close to zero. For example, suppose there are 20 foxes and 50 hares as before, but now we bring in more foxes, bringing their number to 200. If we run the computation, we find the number of hares will plummet to just slightly more than 1 hare in the whole forest. In reality that most likely means the hares die out, and then the foxes will die out as well as they will have nothing to eat.

Showing that a system of equations has a stable solution can be a very difficult problem. When Isaac Newton put forth his laws of planetary motions, he proved that a single planet orbiting a single sun is a stable system. But any solar system with more than 1 planet proved very difficult indeed. In fact, such a system behaves chaotically (see § 3.5), meaning small changes in initial conditions lead to very different long-term outcomes. From numerical experimentation and measurements, we know the earth will not fly out into the empty space or crash into the sun, for at least some millions of years or so. But we do not know what happens beyond that.

3.3.3 Exercises

***Exercise* 3.3.1:** *Take the* damped nonlinear pendulum equation $\theta'' + \mu\theta' + (g/L)\sin\theta = 0$ *for some* $\mu > 0$ *(that is, there is some friction).*

a) Suppose $\mu = 1$ *and* $g/L = 1$ *for simplicity, find and classify the critical points.*

b) Do the same for any $\mu > 0$ *and any* g *and* L*, but such that the damping is small, in particular,* $\mu^2 < 4(g/L)$.

c) Explain what your findings mean, and if it agrees with what you expect in reality.

***Exercise* 3.3.2:** *Suppose the hares do not grow exponentially, but logistically. In particular consider*

$$x' = (0.4 - 0.01y)x - \gamma x^2, \qquad y' = (0.003x - 0.3)y.$$

For the following two values of γ*, find and classify all the critical points in the positive quadrant, that is, for* $x \geq 0$ *and* $y \geq 0$*. Then sketch the phase diagram. Discuss the implication for the long term behavior of the population.*

a) $\gamma = 0.001$, *b)* $\gamma = 0.01$.

***Exercise* 3.3.3:**

a) Suppose x *and* y *are positive variables. Show* $\frac{yx}{e^{x+y}}$ *attains a maximum at* $(1,1)$.

b) Suppose a, b, c, d *are positive constants, and also suppose* x *and* y *are positive variables. Show* $\frac{y^a x^d}{e^{cx+by}}$ *attains a maximum at* $(d/c, a/b)$.

***Exercise* 3.3.4:** *Suppose that for the pendulum equation we take a trajectory giving the spinning-around motion, for example* $\omega = \sqrt{\frac{2g}{L}\cos\theta + \frac{2g}{L} + \omega_0^2}$*. This is the trajectory where the lowest angular velocity is* ω_0^2*. Find an integral expression for how long it takes the pendulum to go all the way around.*

***Exercise* 3.3.5** (challenging)*: Take the pendulum, suppose the initial position is $\theta = 0$.*

a) Find the expression for ω giving the trajectory with initial condition $(0, \omega_0)$. Hint: Figure out what C should be in terms of ω_0.

b) Find the crucial angular velocity ω_1, such that for any higher initial angular velocity, the pendulum will keep going around its axis, and for any lower initial angular velocity, the pendulum will simply swing back and forth. Hint: When the pendulum doesn't go over the top the expression for ω will be undefined for some θs.

c) What do you think happens if the initial condition is $(0, \omega_1)$, that is, the initial angle is 0, and the initial angular velocity is exactly ω_1.

Exercise* 3.3.101: Take the damped nonlinear pendulum equation $\theta'' + \mu\theta' + (g/L)\sin\theta = 0$ for some $\mu > 0$ (that is, there is friction). Suppose the friction is large, in particular $\mu^2 > 4(g/L)$.*

a) Find and classify the critical points.

b) Explain what your findings mean, and if it agrees with what you expect in reality.

Exercise* 3.3.102: Suppose we have the system predator-prey system where the foxes are also killed at a constant rate h (h foxes killed per unit time): $x' = (a - by)x$, $y' = (cx - d)y - h$.*

a) Find the critical points and the Jacobian matrices of the system.

b) Put in the constants $a = 0.4$, $b = 0.01$, $c = 0.003$, $d = 0.3$, $h = 10$. Analyze the critical points. What do you think it says about the forest?

***Exercise* 3.3.103** (challenging)*: Suppose the foxes never die. That is, we have the system $x' = (a - by)x$, $y' = cxy$. Find the critical points and notice they are not isolated. What will happen to the population in the forest if it starts at some positive numbers. Hint: Think of the constant of motion.*

3.4 Limit cycles

Note: less than 1 lecture, discussed in §6.1 and §6.4 in [EP] , §9.7 in [BD]

For nonlinear systems, trajectories do not simply need to approach or leave a single point. They may in fact approach a larger set, such as a circle or another closed curve.

Example 3.4.1: The *Van der Pol oscillator** is the following equation

$$x'' - \mu(1 - x^2)x' + x = 0,$$

where μ is some positive constant. The Van der Pol oscillator originated with electrical circuits, but finds applications in diverse fields such as biology, seismology, and other physical sciences.

For simplicity, let us use $\mu = 1$. A phase diagram is given in the left-hand plot in Figure 3.10. Notice how the trajectories seem to very quickly settle on a closed curve. On the right-hand side is the plot of a single solution for $t = 0$ to $t = 30$ with initial conditions $x(0) = 0.1$ and $x'(0) = 0.1$. The solution quickly tends to a periodic solution.

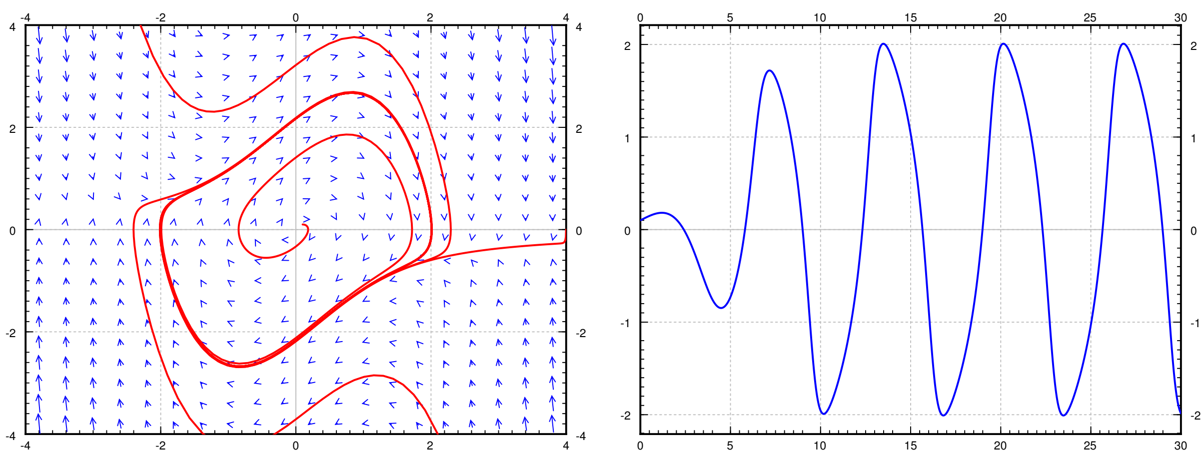


Figure 3.10: The phase portrait (left) and a graph of a sample solution of the Van der Pol oscillator.

The Van der Pol oscillator is an example of so-called *relaxation oscillation*. The word relaxation comes from the sudden jump (the very steep part of the solution). For larger μ the steep part becomes even more pronounced, for small μ the limit cycle looks more like a circle. In fact, setting $\mu = 0$, we get $x'' + x = 0$, which is a linear system with a center and all trajectories become circles.

A trajectory in the phase portrait that is a closed curve (a curve that is a loop) is called a *closed trajectory*. A *limit cycle* is a closed trajectory such that at least one other trajectory spirals into it (or spirals out of it). For example, the closed curve in the phase portrait for

*Named for the Dutch physicist Balthasar van der Pol (1889–1959).

the Van der Pol equation is a limit cycle. If all trajectories that start near the limit cycle spiral into it, the limit cycle is called *asymptotically stable*. The limit cycle in the Van der Pol oscillator is asymptotically stable.

Given a closed trajectory on an autonomous system, any solution that starts on it is periodic. Such a curve is called a *periodic orbit*. More precisely, if $\big(x(t), y(t)\big)$ is a solution such that for some t_0 the point $\big(x(t_0), y(t_0)\big)$ lies on a periodic orbit, then both $x(t)$ and $y(t)$ are periodic functions (with the same period). That is, there is some number P such that $x(t) = x(t+P)$ and $y(t) = y(t+P)$.

Consider the system

$$x' = f(x, y), \qquad y' = g(x, y), \tag{3.2}$$

where the functions f and g have continuous derivatives in some region R in the plane.

Theorem 3.4.1 (Poincaré–Bendixson*)**.** *Suppose R is a closed bounded region (a region in the plane that includes its boundary and does not have points arbitrarily far from the origin). Suppose $\big(x(t), y(t)\big)$ is a solution of (3.2) in R that exists for all $t \geq t_0$. Then either the solution is a periodic function, or the solution tends towards a periodic solution in R.*

The main point of the theorem is that if you find one solution that exists for all t large enough (that is, as t goes to infinity) and stays within a bounded region, then you have found either a periodic orbit, or a solution that spirals towards a limit cycle or tends to a critical point. That is, in the long term, the behavior is very close to a periodic function. Note that a constant solution at a critical point is periodic (with any period). The theorem is more a qualitative statement rather than something to help us in computations. In practice it is hard to find analytic solutions and so hard to show rigorously that they exist for all time. But if we think the solution exists we numerically solve for a large time to approximate the limit cycle. Another caveat is that the theorem only works in two dimensions. In three dimensions and higher, there is simply too much room.

The theorem applies to all solutions in the Van der Pol oscillator. Solutions that start at any point except the origin $(0, 0)$ tend to the periodic solution around the limit cycle, and the initial condition of $(0, 0)$ gives the constant solution $x = 0$, $y = 0$.

Example 3.4.2: Consider

$$x' = y + (x^2 + y^2 - 1)^2 x, \qquad y' = -x + (x^2 + y^2 - 1)^2 y.$$

A vector field along with solutions with initial conditions $(1.02, 0)$, $(0.9, 0)$, and $(0.1, 0)$ are drawn in Figure 3.11 on the next page.

Notice that points on the unit circle (distance one from the origin) satisfy $x^2 + y^2 - 1 = 0$. And $x(t) = \sin(t)$, $y = \cos(t)$ is a solution of the system. Therefore we have a closed trajectory. For points off the unit circle, the second term in x' pushes the solution further away from the y-axis than the system $x' = y$, $y' = -x$, and y' pushes the solution further away from the x-axis than the linear system $x' = y$, $y' = -x$. In other words for all other initial conditions the trajectory will spiral out.

*Ivar Otto Bendixson (1861–1935) was a Swedish mathematician.

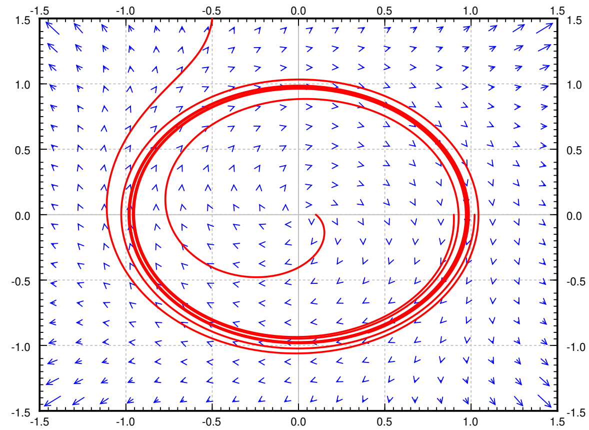


Figure 3.11: Unstable limit cycle example.

This means that for initial conditions inside the unit circle, the solution spirals out towards the periodic solution on the unit circle, and for initial conditions outside the unit circle the solutions spiral off towards infinity. Therefore the unit circle is a limit cycle, but not an asymptotically stable one. The Poincaré–Bendixson Theorem applies to the initial points inside the unit circle, as those solutions stay bounded, but not to those outside, as those solutions go off to infinity.

A very similar analysis applies to the system

$$x' = y + (x^2 + y^2 - 1)x, \qquad y' = -x + (x^2 + y^2 - 1)y.$$

We still obtain a closed trajectory on the unit circle, and points outside the unit circle spiral out to infinity, but now points inside the unit circle spiral towards the critical point at the origin. So this system does not have a limit cycle, even though it has a closed trajectory.

Due to the Picard theorem (Theorem 2.1.1 on page 77) we find that no matter where we are in the plane we can always find a solution a little bit further in time, as long as f and g have continuous derivatives. So if we find a closed trajectory in an autonomous system, then for every initial point inside the closed trajectory, the solution will exist for all time and it will stay bounded (it will stay inside the closed trajectory). So the moment we found the solution above going around the unit circle, we knew that for every initial point inside the circle, the solution exists for all time and the Poincaré–Bendixson theorem applies.

Let us next look for conditions when limit cycles (or periodic orbits) do not exist. We assume the equation (3.2) is defined on a *simply connected region*, that is, a region with no holes we can go around. For example the entire plane is a simply connected region, and so is the inside of the unit disc. However, the entire plane minus a point is not a simply connected region as it has a "hole" at the origin.

Theorem 3.4.2 (Bendixson–Dulac*). *Suppose R is a simply connected region, and the expression*†

$$\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}$$

is either always positive or always negative on R (except perhaps a small set such as on isolated points or curves) then the system (3.2) has no closed trajectory inside R.

The theorem gives us a way of ruling out the existence of a closed trajectory, and hence a way of ruling out limit cycles. The exception about points or curves means that we can allow the expression to be zero at a few points, or perhaps on a curve, but not on any larger set.

Example 3.4.3: Let us look at $x' = y + y^2 e^x$, $y' = x$ in the entire plane (see Example 3.2.2 on page 133). The entire plane is simply connected and so we can apply the theorem. We compute $\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = y^2 e^x + 0$. The function $y^2 e^x$ is always positive except on the line $y = 0$. Therefore, via the theorem, the system has no closed trajectories.

In some books (or the internet) the theorem is not stated carefully and it concludes there are no periodic solutions. That is not quite right. The example above has two critical points and hence it has constant solutions, and constant functions are periodic. The conclusion of the theorem should be that there exist no trajectories that form closed curves. Another way to state the conclusion of the theorem would be to say that there exist no nonconstant periodic solutions that stay in R.

Example 3.4.4: Let us look at a somewhat more complicated example. Take the system $x' = -y - x^2$, $y' = -x + y^2$ (see Example 3.2.1 on page 132). We compute $\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = -2x + 2y = 2(-x + y)$. This expression takes on both signs, so if we are talking about the whole plane we cannot simply apply the theorem. However, we could apply it on the set where $-x + y \geq 0$. Via the theorem, there is no closed trajectory in that set. Similarly, there is no closed trajectory in the set $-x + y \leq 0$. We cannot conclude (yet) that there is no closed trajectory in the entire plane. For all we know, perhaps half of it is in the set where $-x + y \geq 0$ and the other half is in the set where $-x + y \leq 0$.

The key is to look at the line where $-x+y = 0$, or $x = y$. On this line $x' = -y-x^2 = -x-x^2$ and $y' = -x + y^2 = -x + x^2$. In particular, when $x = y$ then $x' \leq y'$. That means that the arrows, the vectors (x', y'), always point into the set where $-x + y \geq 0$. There is no way we can start in the set where $-x + y \geq 0$ and go into the set where $-x + y \leq 0$. Once we are in the set where $-x + y \geq 0$, we stay there. So no closed trajectory can have points in both sets.

Example 3.4.5: Consider $x' = y + (x^2 + y^2 - 1)x$, $y' = -x + (x^2 + y^2 - 1)y$, and consider the region R given by $x^2 + y^2 > \frac{1}{2}$. That is, R is the region outside a circle of radius $\frac{1}{\sqrt{2}}$

*Henri Dulac (1870–1955) was a French mathematician.

†Usually the expression in the Bendixson–Dulac Theorem is $\frac{\partial(\varphi f)}{\partial x} + \frac{\partial(\varphi g)}{\partial y}$ for some continuously differentiable function φ. For simplicity, let us just consider the case $\varphi = 1$.

centered at the origin. Then there is a closed trajectory in R, namely $x = \cos(t)$, $y = \sin(t)$. Furthermore,

$$\frac{\partial f}{\partial x} + \frac{\partial g}{\partial x} = 4x^2 + 4y^2 - 2,$$

which is always positive on R. So what is going on? The Bendixson–Dulac theorem does not apply since the region R is not simply connected—it has a hole, the circle we cut out!

3.4.1 Exercises

Exercise* 3.4.1*: *Show that the following systems have no closed trajectories.*

a) $x' = x^3 + y, \quad y' = y^3 + x^2,$ *b)* $x' = e^{x-y}, \quad y' = e^{x+y},$

c) $x' = x + 3y^2 - y^3, \quad y' = y^3 + x^2.$

Exercise* 3.4.2*: *Formulate a condition for a 2-by-2 linear system $\vec{x}' = A\vec{x}$ to not be a center using the Bendixson–Dulac theorem. That is, the theorem says something about certain elements of A.*

Exercise* 3.4.3*: *Explain why the Bendixson–Dulac Theorem does not apply for any conservative system $x'' + h(x) = 0$.*

Exercise* 3.4.4*: *A system such as $x' = x, y' = y$ has solutions that exist for all time t, yet there are no closed trajectories. Explain why the Poincaré–Bendixson Theorem does not apply.*

Exercise* 3.4.5*: *Differential equations can also be given in different coordinate systems. Suppose we have the system $r' = 1 - r^2$, $\theta' = 1$ given in polar coordinates. Find all the closed trajectories and check if they are limit cycles and if so, if they are asymptotically stable or not.*

Exercise* 3.4.101*: *Show that the following systems have no closed trajectories.*

a) $x' = x + y^2, \quad y' = y + x^2,$ *b)* $x' = -x\sin^2(y), \quad y' = e^x,$

c) $x' = xy^2, \quad y' = x + x^2.$

Exercise* 3.4.102*: *Suppose an autonomous system in the plane has a solution $x = \cos(t) + e^{-t}$, $y = \sin(t) + e^{-t}$. What can you say about the system (in particular about limit cycles and periodic solutions)?*

Exercise* 3.4.103*: *Show that the limit cycle of the Van der Pol oscillator (for $\mu > 0$) must not lie completely in the set where $-1 < x < 1$. Compare with Figure 3.10 on page 147.*

Exercise* 3.4.104*: *Suppose we have the system $r' = \sin(r)$, $\theta' = 1$ given in polar coordinates. Find all the closed trajectories.*

3.5 Chaos

Note: 1 lecture, §6.5 in [EP], §9.8 in [BD]

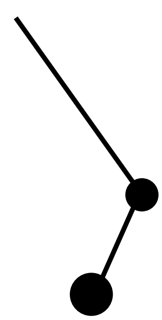
You have surely heard the story about the flap of a butterfly wing in the Amazon causing hurricanes in the North Atlantic. In a prior section, we mentioned that a small change in initial conditions of the planets can lead to very different configuration of the planets in the long term. These are examples of *chaotic systems*. Mathematical chaos is not really chaos, there is precise order behind the scenes. Everything is still deterministic. However a chaotic system is extremely sensitive to initial conditions. This also means even small errors induced via numerical approximation create large errors very quickly, so it is almost impossible to numerically approximate for long times. This is a large part of the trouble, as chaotic systems cannot be in general solved analytically.

Take the weather, the most well-known chaotic system. A small change in the initial conditions (the temperature at every point of the atmosphere for example) produces drastically different predictions in relatively short time, and so we cannot accurately predict weather. And we do not actually know the exact initial conditions. We measure temperatures at a few points with some error, and then we somehow estimate what is in between. There is no way we can accurately measure the effects of every butterfly wing. Then we solve the equations numerically introducing new errors. You should not trust weather prediction more than a few days out.

Chaotic behavior was first noticed by Edward Lorenz* in the 1960s when trying to model thermally induced air convection (movement). Lorentz was looking at the relatively simple system:

$$x' = -10x + 10y, \qquad y' = 28x - y - xz, \qquad z' = -\frac{8}{3}z + xy.$$

A small change in the initial conditions yields a very different solution after a reasonably short time.



A simple example the reader can experiment with, and which displays chaotic behavior, is a double pendulum. The equations for this setup are somewhat complicated, and their derivation is quite tedious, so we will not bother to write them down. The idea is to put a pendulum on the end of another pendulum. The movement of the bottom mass will appear chaotic. This type of chaotic system is a basis for a whole number of office novelty desk toys. It is simple to build a version. Take a piece of a string. Tie two heavy nuts at different points of the string; one at the end, and one a bit above. Now give the bottom nut a little push. As long as the swings are not too big and the string stays tight, you have a double pendulum system.

*Edward Norton Lorenz (1917–2008) was an American mathematician and meteorologist.

3.5.1 Duffing equation and strange attractors

Let us study the so-called *Duffing equation*:

$$x'' + ax' + bx + cx^3 = C\cos(\omega t).$$

Here a, b, c, C, and ω are constants. Except for the cx^3 term, this equation looks like a forced mass-spring system. The cx^3 means the spring does not exactly obey Hooke's law (which no real-world spring actually obeys exactly). When c is not zero, the equation does not have a closed form solution, so we must resort to numerical solutions, as is usual for nonlinear systems. Not all choices of constants and initial conditions exhibit chaotic behavior. Let us study

$$x'' + 0.05x' + x^3 = 8\cos(t).$$

The equation is not autonomous, so we cannot draw the vector field in the phase plane. We can still draw trajectories. In Figure 3.12, we plot trajectories for t going from 0 to 15 for two very close initial conditions $(2, 3)$ and $(2, 2.9)$, and also the solutions in the (x, t) space. The two trajectories are close at first, but after a while diverge significantly. This sensitivity to initial conditions is precisely what we mean by the system behaving chaotically.

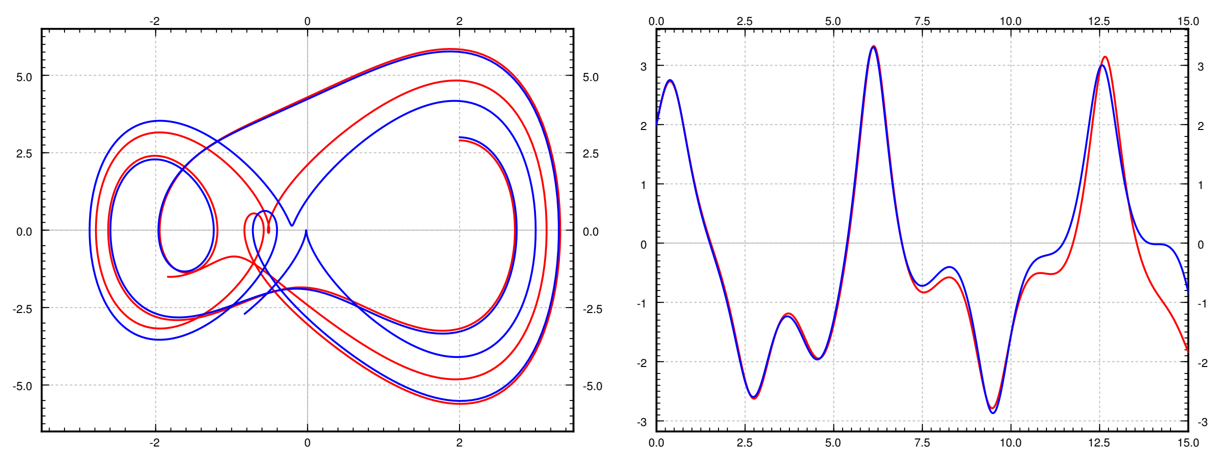


Figure 3.12: On left, two trajectories in phase space for $0 \leq t \leq 15$, *for the Duffing equation one with initial conditions* $(2, 3)$ *and the other with* $(2, 2.9)$. *On right the two solutions in* (x, t)*-space.*

Let us see the long term behavior. In Figure 3.13 on the next page, we plot the behavior of the system for initial conditions $(2, 3)$ for a longer period of time. It is hard to see any particular pattern in the shape of the solution except that it seems to oscillate, but each oscillation appears quite unique. The oscillation is expected due to the forcing term. We mention that to produce the picture accurately, a ridiculously large number of steps* had to be used in the numerical algorithm, as even small errors quickly propagate in a chaotic system.

*In fact for reference, 30,000 steps were used with the Runge–Kutta algorithm, see exercises in § 1.7.

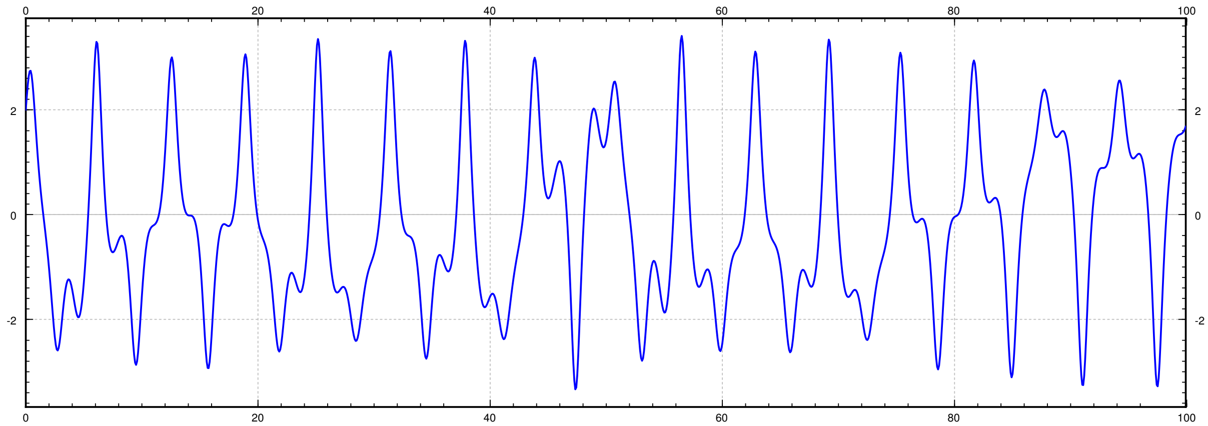


Figure 3.13: The solution to the given Duffing equation for t from 0 to 100.

It is very difficult to analyze chaotic systems, or to find the order behind the madness, but let us try to do something that we did for the standard mass-spring system. One way we analyzed the system is that we figured out what was the long term behavior (not dependent on initial conditions). From the figure above, it is clear that we will not get a nice exact description of the long term behavior for this chaotic system, but perhaps we can find some order to what happens on each "oscillation" and what do these oscillations have in common.

The concept we explore is that of a *Poincaré section**. Instead of looking at t in a certain interval, we look at where the system is at a certain sequence of points in time. Imagine flashing a strobe at a fixed frequency and drawing the points where the solution is during the flashes. The right strobing frequency depends on the system in question. The correct frequency for the forced Duffing equation (and other similar systems) is the frequency of the forcing term. For the Duffing equation above, find a solution $\big(x(t), y(t)\big)$, and look at the points

$$\big(x(0), y(0)\big), \quad \big(x(2\pi), y(2\pi)\big), \quad \big(x(4\pi), y(4\pi)\big), \quad \big(x(6\pi), y(6\pi)\big), \quad \ldots$$

As we are really not interested in the transient part of the solution, that is, the part of the solution that depends on the initial condition, we skip some number of steps in the beginning. For example, we might skip the first 100 such steps and start plotting points at $t = 100(2\pi)$, that is

$$\big(x(200\pi), y(200\pi)\big), \quad \big(x(202\pi), y(202\pi)\big), \quad \big(x(204\pi), y(204\pi)\big), \quad \ldots$$

The plot of these points is the Poincaré section. After plotting enough points, a curious pattern emerges in Figure 3.14 on the facing page (the left-hand picture), a so-called *strange attractor*.

*Named for the French polymath Jules Henri Poincaré (1854–1912).

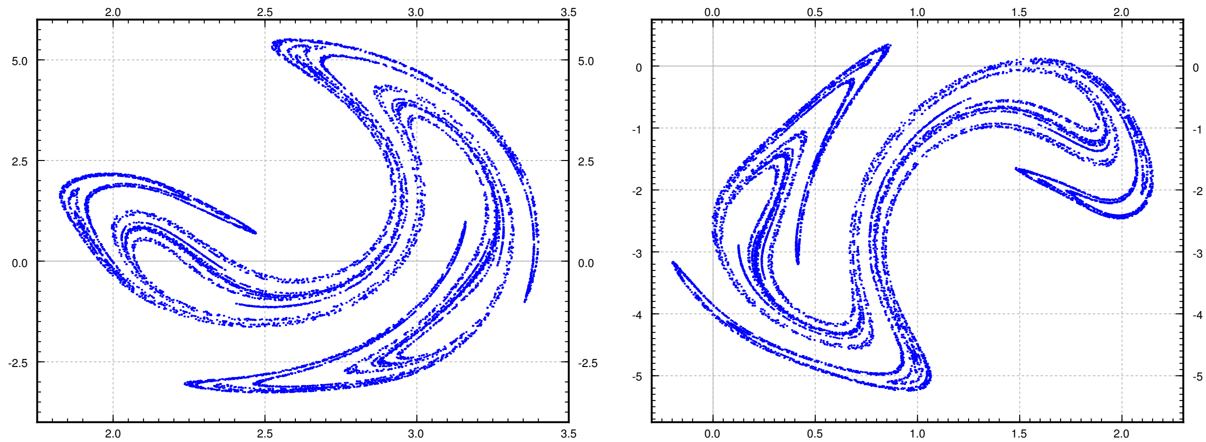


Figure 3.14: Strange attractor. The left plot is with no phase shift, the right plot has phase shift $\pi/4$.

Given a sequence of points, an *attractor* is a set towards which the points in the sequence eventually get closer and closer to, that is, they are attracted. The Poincaré section is not really the attractor itself, but as the points are very close to it, we see its shape. The strange attractor is a very complicated set. It has fractal structure, that is, if you zoom in as far as you want, you keep seeing the same complicated structure.

The initial condition makes no difference. If we start with a different initial condition, the points eventually gravitate towards the attractor, and so as long as we throw away the first few points, we get the same picture. Similarly small errors in the numerical approximations do not matter here.

An amazing thing is that a chaotic system such as the Duffing equation is not random at all. There is a very complicated order to it, and the strange attractor says something about this order. We cannot quite say what state the system will be in eventually, but given the fixed strobing frequency we narrow it down to the points on the attractor.

If we use a phase shift, for example $\pi/4$, and look at the times

$$\pi/4, \quad 2\pi + \pi/4, \quad 4\pi + \pi/4, \quad 6\pi + \pi/4, \quad \ldots$$

we obtain a slightly different attractor. The picture is the right-hand side of Figure 3.14. It is as if we had rotated, moved, and slightly distorted the original. For each phase shift you can find the set of points towards which the system periodically keeps coming back to.

Study the pictures and notice especially the scales—where are these attractors located in the phase plane. Notice the regions where the strange attractor lives and compare it to the plot of the trajectories in Figure 3.12 on page 153.

Let us compare this section to the discussion in § **??** on forced oscillations. Consider

$$x'' + 2px' + \omega_0^2 x = \frac{F_0}{m}\cos(\omega t).$$

This is like the Duffing equation, but with no x^3 term. The steady periodic solution is of the form

$$x = C\cos(\omega t + \gamma).$$

Strobing using the frequency ω, we obtain a single point in the phase space. The attractor in this setting is a single point—an expected result as the system is not chaotic. It was the opposite of chaotic: Any difference induced by the initial conditions dies away very quickly, and we settle into always the same steady periodic motion.

3.5.2 The Lorenz system

In two dimensions to find chaotic behavior, we must study forced, or non-autonomous, systems such as the Duffing equation. The Poincaré–Bendixson Theorem says that a solution to an autonomous two-dimensional system that exists for all time in the future and does not go towards infinity is periodic or tends towards a periodic solution. Hardly the chaotic behavior we are looking for.

In three dimensions, even autonomous systems can be chaotic. Let us very briefly return to the Lorenz system

$$x' = -10x + 10y, \qquad y' = 28x - y - xz, \qquad z' = -\frac{8}{3}z + xy.$$

The Lorenz system is an autonomous system in three dimensions exhibiting chaotic behavior. See the Figure 3.15 for a sample trajectory, which is now a curve in three-dimensional space.

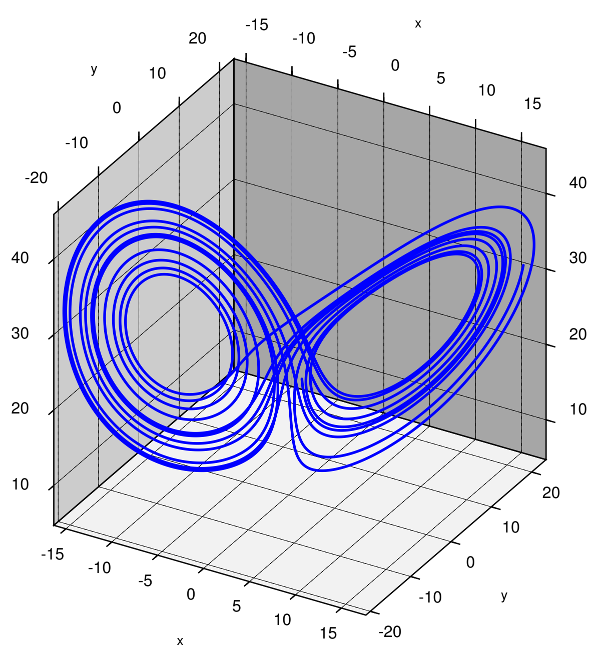


Figure 3.15: A trajectory in the Lorenz system.

The solutions tend to an *attractor* in space, the so-called *Lorenz attractor*. In this case no strobing is necessary, the solution will tend towards the attractor set. Again we cannot quite see the attractor itself, but if we try to follow a solution for long enough, as in the figure, we get a pretty good picture of what the attractor looks like. The Lorenz attractor is also a strange attractor and has a complicated fractal structure. And, just as for the Duffing equation, what we want to draw is not the whole trajectory, but start drawing the trajectory after a while, once it is close to the attractor.

The path of the trajectory is not simply a repeating figure-eight. The trajectory spins some seemingly random number of times on the left, then spins a number of times on the right, and so on. As this system arose in weather prediction, one can perhaps imagine a few days of warm weather and then a few days of cold weather, where it is not easy to predict when the weather will change, just as it is not really easy to predict far in advance when the solution will jump onto the other side. See Figure 3.16 for a plot of the x component of the solution drawn above. A negative x corresponds to the left "loop" and a positive x corresponds to the right "loop".

Most of the mathematics we studied in this book is quite classical and well understood. On the other hand, chaos, including the Lorenz system, continues to be the subject of current research. Furthermore, chaos has found applications not just in the sciences, but also in art.

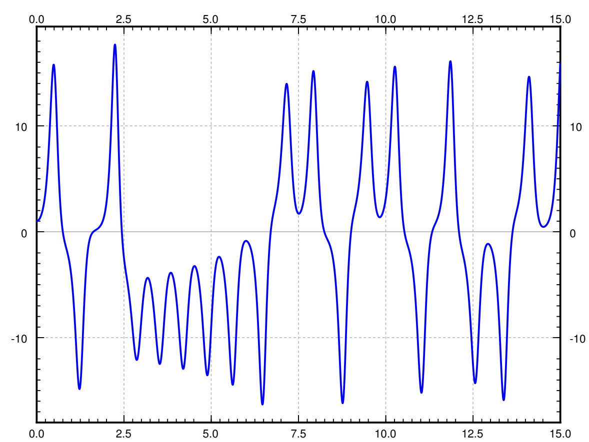


Figure 3.16: Graph of the $x(t)$ component of the solution.

3.5.3 Exercises

Exercise* 3.5.1*: *For the non-chaotic equation $x'' + 2px' + \omega_0^2 x = \frac{F_0}{m}\cos(\omega t)$, suppose we strobe with frequency ω as we mentioned above. Use the known steady periodic solution to find precisely the point which is the attractor for the Poincaré section.*

Exercise* 3.5.2** (project):*** *A simple fractal attractor can be drawn via the following chaos game. Draw the three vertices of a triangle and label them, say p_1, p_2 and p_3. Draw some random point p (it does not have to be one of the three points above). Roll a die to pick of the p_1, p_2, or p_3 randomly (for example 1 and 4 mean p_1, 2 and 5 mean p_2, and 3 and 6 mean p_3). Suppose we picked p_2, then let p_{new} be the point exactly halfway between p and p_2. Draw this point and let p now refer to this new point p_{new}. Rinse, repeat. Try to be precise and draw as many iterations as possible. Your points will be attracted to the so-called* Sierpinski triangle*. A computer was used to run the game for 10,000 iterations to obtain the picture in Figure 3.17.*

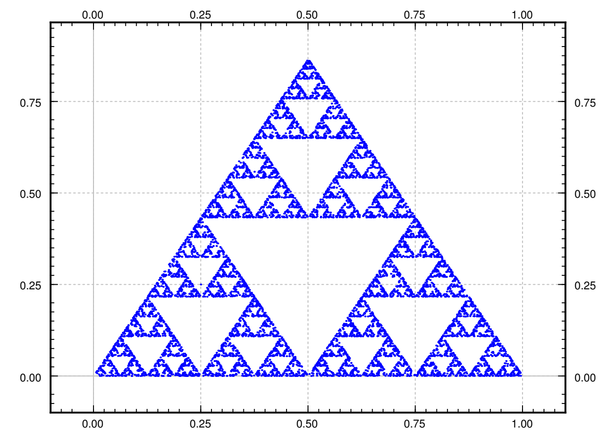


Figure 3.17: 10,000 iterations of the chaos game producing the Sierpinski triangle.

Exercise* 3.5.3** (project):*** *Construct the double pendulum described in the text with a string and two nuts (or heavy beads). Play around with the position of the middle nut, and perhaps use different weight nuts. Describe what you find.*

Exercise* 3.5.4** (computer project):*** *Use a computer software (such as Matlab, Octave, or perhaps even a spreadsheet), plot the solution of the given forced Duffing equation with Euler's method. Plotting the solution for t from 0 to 100 with several different (small) step sizes. Discuss.*

Exercise* 3.5.101*: *Find critical points of the Lorenz system and the associated linearizations.*

Appendix A

Linear algebra

A.1 Vectors, mappings, and matrices

Note: 2 lectures

In real life, there is most often more than one variable. We wish to organize dealing with multiple variables in a consistent manner, and in particular organize dealing with linear equations and linear mappings, as those are both rather useful and rather easy to handle. Mathematicians joke that "to an engineer every problem is linear, and everything is a matrix." And well, they (the engineers) are not wrong. Quite often, solving an engineering problem is figuring out the right finite-dimensional linear problem to solve, which is then solved with some matrix manipulation. Most importantly, linear problems are the ones that we know how to solve, and we have many tools to solve them. For engineers, mathematicians, physicists, and anybody else in a technical field, it is absolutely vital to learn linear algebra.

As motivation, suppose we wish to solve

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

for x and y. That is, we desire numbers x and y such that the two equations are satisfied. Let us perhaps start by adding the equations together to find

$$
x + 2x - y + y = 2 + 4, \qquad \text{or} \qquad 3x = 6.
$$

In other words, $x = 2$. Once we have that, we plug $x = 2$ into the first equation to find $2 - y = 2$, so $y = 0$. OK, that was easy. What is all this fuss about linear equations. Well, try doing this if you have 5000 unknowns[†]. Also, we may have such equations not just of numbers, but of functions and derivatives of functions in differential equations. Clearly we need a systematic way of doing things. A nice consequence of making things systematic

[†]One of the downsides of making everything look like a linear problem is that the number of variables tends to become huge.

and simpler to write down is that it becomes easier to have computers do the work for us. Computers are rather stupid, they do not think, but are very good at doing lots of repetitive tasks precisely, as long as we figure out a systematic way for them to perform the tasks.

A.1.1 Vectors and operations on vectors

Consider n real numbers as an n-tuple:

$$(x_1, x_2, \ldots, x_n).$$

The set of such n-tuples is the so-called *n-dimensional space*, often denoted by $\mathbb{R}^n$. Sometimes we call this the n-dimensional *euclidean space**. In two dimensions, $\mathbb{R}^2$ is called the *cartesian plane*†. Each such n-tuple represents a point in the n-dimensional space. For example, the point $(1, 2)$ in the plane $\mathbb{R}^2$ is one unit to the right and two units up from the origin.

When we do algebra with these n-tuples of numbers we call them *vectors*‡. Mathematicians are keen on separating what is a vector and what is a point of the space or in the plane, and it turns out to be an important distinction, however, for the purposes of linear algebra we can think of everything being represented by a vector. A way to think of a vector, which is especially useful in calculus and differential equations, is an arrow. It is an object that has a *direction* and a *magnitude*. For instance, the vector $(1, 2)$ is the arrow from the origin to the point $(1, 2)$ in the plane. The magnitude is the length of the arrow. See Figure A.1. If we think of vectors as arrows, the arrow does not always have to start at the origin. If we do move it around, however, it should always keep the same direction and the same magnitude.

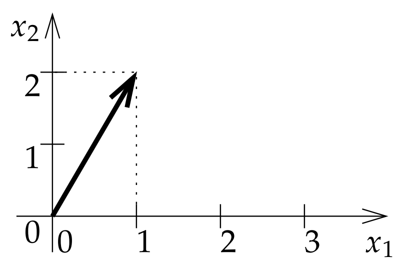


Figure A.1: The vector $(1, 2)$ *drawn as an arrow from the origin to the point* $(1, 2)$.

As vectors are arrows, when we want to give a name to a vector, we draw a little arrow above it:

$$\vec{x}$$

*Named after the ancient Greek mathematician Euclid of Alexandria (around 300 BC), possibly the most famous of mathematicians; even small towns often have Euclid Street or Euclid Avenue.

†Named after the French mathematician René Descartes (1596–1650). It is "cartesian" as his name in Latin is Renatus Cartesius.

‡A common notation to distinguish vectors from points is to write $(1, 2)$ for the point and $\langle 1, 2 \rangle$ for the vector. We write both as $(1, 2)$.

Another popular notation is a bold $\mathbf{x}$, although we will use the little arrows. It may be easy to write a bold letter in a book, but it is not so easy to write it by hand on paper or on the board. Mathematicians often do not even write the arrows. A mathematician would write x and remember that x is a vector and not a number. Just like you remember that Bob is your uncle, and you don't have to keep repeating "Uncle Bob" and you can just say "Bob." In this book, however, we will call Bob "Uncle Bob" and write vectors with the little arrows.

The *magnitude* can be computed using the Pythagorean theorem. The vector $(1,2)$ drawn in the figure has magnitude $\sqrt{1^2+2^2} = \sqrt{5}$. The magnitude is denoted by $\|\vec{x}\|$, and, in any number of dimensions, it can be computed in the same way:

$$\|\vec{x}\| = \|(x_1, x_2, \dots, x_n)\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

For reasons that will become clear in the next section, we often write vectors as so-called *column vectors*:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Don't worry. It is just a different way of writing the same thing. For example, the vector $(1,2)$ can be written as

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The fact that we write arrows above vectors allows us to write several vectors $\vec{x}_1$, $\vec{x}_2$, etc., without confusing these with the components of some other vector $\vec{x}$.

So where is the *algebra* from *linear algebra*? Well, arrows can be added, subtracted, and multiplied by numbers. First we consider *addition*. If we have two arrows, we simply move along one, and then along the other. See Figure A.2.

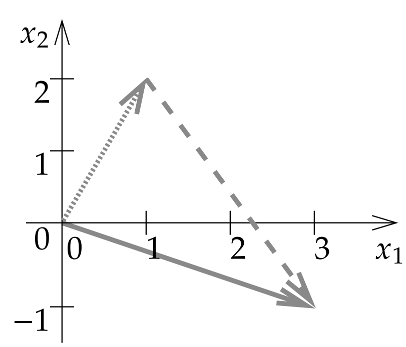


Figure A.2: Adding the vectors $(1,2)$, *drawn dotted, and* $(2,-3)$, *drawn dashed. The result,* $(3,-1)$, *is drawn as a solid arrow.*

It is rather easy to see what it does to the numbers that represent the vectors. Suppose we want to add $(1,2)$ to $(2,-3)$ as in the figure. We travel along $(1,2)$ and then we travel along $(2,-3)$. What we did was travel one unit right, two units up, and then we travelled two units right, and three units down (the negative three). That means that we ended up at $\bigl(1+2, 2+(-3)\bigr) = (3,-1)$. And that's how addition always works:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Subtracting is similar. What $\vec{x} - \vec{y}$ means visually is that we first travel along $\vec{x}$, and then we travel backwards along $\vec{y}$. See Figure A.3. It is like adding $\vec{x} + (-\vec{y})$ where $-\vec{y}$ is the arrow we obtain by erasing the arrow head from one side and drawing it on the other side, that is, we reverse the direction. In terms of the numbers, we simply go backwards both horizontally and vertically, so we negate both numbers. For instance, if $\vec{y}$ is $(-2,1)$, then $-\vec{y}$ is $(2,-1)$.

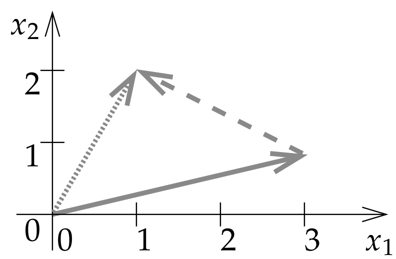


Figure A.3: Subtraction, the vector $(1,2)$*, drawn dotted, minus* $(-2,1)$*, drawn dashed. The result,* $(3,1)$*, is drawn as a solid arrow.*

Another intuitive thing to do to a vector is to *scale* it. We represent this by multiplication of a number with a vector. Because of this, when we wish to distinguish between vectors and numbers, we call the numbers *scalars*. For example, suppose we want to travel three times further. If the vector is $(1,2)$, travelling 3 times further means going 3 units to the right and 6 units up, so we get the vector $(3,6)$. We just multiply each number in the vector by 3. If α is a number, then

$$\alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}.$$

Scaling (by a positive number) multiplies the magnitude and leaves direction untouched. The magnitude of $(1,2)$ is $\sqrt{5}$. The magnitude of 3 times $(1,2)$, that is, $(3,6)$, is $3\sqrt{5}$.

When the scalar is negative, then when we multiply a vector by it, the vector is not only scaled, but it also switches direction. Multiplying $(1,2)$ by -3 means we should go 3 times further but in the opposite direction, so 3 units to the left and 6 units down, or in other words, $(-3,-6)$. As we mentioned above, $-\vec{y}$ is a reverse of $\vec{y}$, and this is the same as $(-1)\vec{y}$.

In Figure A.4, you can see a couple of examples of what scaling a vector means visually.

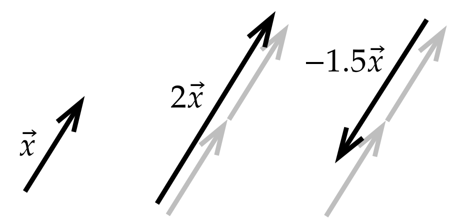


Figure A.4: A vector $\vec{x}$, the vector $2\vec{x}$ (same direction, double the magnitude), and the vector $-1.5\vec{x}$ (opposite direction, 1.5 times the magnitude).

We put all of these operations together to work out more complicated expressions. Let us compute a small example:

$$3\begin{bmatrix}1\\2\end{bmatrix}+2\begin{bmatrix}-4\\-1\end{bmatrix}-3\begin{bmatrix}-2\\2\end{bmatrix}=\begin{bmatrix}3(1)+2(-4)-3(-2)\\3(2)+2(-1)-3(2)\end{bmatrix}=\begin{bmatrix}1\\-2\end{bmatrix}.$$

As we said a vector is a direction and a magnitude. Magnitude is easy to represent, it is just a number. The *direction* is usually given by a vector with magnitude one. We call such a vector a *unit vector*. That is, $\vec{u}$ is a unit vector when $\lVert\vec{u}\rVert = 1$. For instance, the vectors $(1,0)$, $({}^1\!/\!\sqrt{2}, {}^1\!/\!\sqrt{2})$, and $(0,-1)$ are all unit vectors.

To represent the direction of a vector $\vec{x}$, we need to find the unit vector in the same direction. To do so, we simply rescale $\vec{x}$ by the reciprocal of the magnitude, that is $\frac{1}{\lVert\vec{x}\rVert}\vec{x}$, or more concisely $\frac{\vec{x}}{\lVert\vec{x}\rVert}$.

As an example, the unit vector in the direction of $(1,2)$ is the vector

$$\frac{1}{\sqrt{1^2+2^2}}(1,2)=\left(\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}\right).$$

A.1.2 Linear mappings and matrices

A *vector-valued function* F is a rule that takes a vector $\vec{x}$ and returns another vector $\vec{y}$. For example, F could be a scaling that doubles the size of vectors:

$$F(\vec{x}) = 2\vec{x}.$$

Applied to say $(1,3)$ we get

$$F\left(\begin{bmatrix}1\\3\end{bmatrix}\right)=2\begin{bmatrix}1\\3\end{bmatrix}=\begin{bmatrix}2\\6\end{bmatrix}.$$

If F is a mapping that takes vectors in $\mathbb{R}^2$ to $\mathbb{R}^2$ (such as the above), we write

$$F\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

The words *function* and *mapping* are used rather interchangeably, although more often than not, *mapping* is used when talking about a vector-valued function, and the word *function* is often used when the function is scalar-valued.

A beginning student of mathematics (and many a seasoned mathematician), that sees an expression such as

$$f(3x + 8y)$$

yearns to write

$$3f(x) + 8f(y).$$

After all, who hasn't wanted to write $\sqrt{x+y} = \sqrt{x} + \sqrt{y}$ or something like that at some point in their mathematical lives. Wouldn't life be simple if we could do that? Of course we cannot always do that (for example, not with the square roots!) But there are many other functions where we can do exactly the above. Such functions are called *linear*.

A mapping $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is called *linear* if

$$F(\vec{x} + \vec{y}) = F(\vec{x}) + F(\vec{y}),$$

for any vectors $\vec{x}$ and $\vec{y}$, and also

$$F(\alpha\vec{x}) = \alpha F(\vec{x}),$$

for any scalar α. The F we defined above that doubles the size of all vectors is linear. Let us check:

$$F(\vec{x} + \vec{y}) = 2(\vec{x} + \vec{y}) = 2\vec{x} + 2\vec{y} = F(\vec{x}) + F(\vec{y}),$$

and also

$$F(\alpha\vec{x}) = 2\alpha\vec{x} = \alpha 2\vec{x} = \alpha F(\vec{x}).$$

We also call a linear function a *linear transformation*. If you want to be really fancy and impress your friends, you can call it a *linear operator*. When a mapping is linear we often do not write the parentheses. We write simply

$$F\vec{x}$$

instead of $F(\vec{x})$. We do this because linearity means that the mapping F behaves like multiplying $\vec{x}$ by "something." That something is a matrix.

A *matrix* is an $m \times n$ array of numbers (m rows and n columns). A 3×5 matrix is

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix}.$$

The numbers a_{ij} are called *elements* or *entries*.

A column vector is simply an $m \times 1$ matrix. Similarly to a column vector there is also a *row vector*, which is a $1 \times n$ matrix. If we have an $n \times n$ matrix, then we say that it is a *square matrix*.

Now how does a matrix A relate to a linear mapping? Well a matrix tells you where certain special vectors go. Let's give a name to those certain vectors. The *standard basis vectors* of $\mathbb{R}^n$ are

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \cdots, \qquad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

In $\mathbb{R}^3$ these vectors are

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

You may recall from calculus of several variables that these are sometimes called $\vec{\imath}$, $\vec{\jmath}$, $\vec{k}$.

The reason these are called a *basis* is that every other vector can be written as a *linear combination* of them. For example, in $\mathbb{R}^3$ the vector $(4, 5, 6)$ can be written as

$$4\vec{e}_1 + 5\vec{e}_2 + 6\vec{e}_3 = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

So how does a matrix represent a linear mapping? Well, the columns of the matrix are the vectors where A as a linear mapping takes $\vec{e}_1$, $\vec{e}_2$, etc. For instance, consider

$$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

As a linear mapping $M \colon \mathbb{R}^2 \to \mathbb{R}^2$ takes $\vec{e}_1 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$ and $\vec{e}_2 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right]$. In other words,

$$M\vec{e}_1 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \text{and} \qquad M\vec{e}_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

More generally, if we have an $n \times m$ matrix A, that is, we have n rows and m columns, then the mapping $A \colon \mathbb{R}^m \to \mathbb{R}^n$ takes $\vec{e}_j$ to the j^{th} column of A. For example,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix}$$

represents a mapping from $\mathbb{R}^5$ to $\mathbb{R}^3$ that does

$$A\vec{e}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}, \qquad A\vec{e}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}, \qquad A\vec{e}_3 = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}, \qquad A\vec{e}_4 = \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix}, \qquad A\vec{e}_5 = \begin{bmatrix} a_{15} \\ a_{25} \\ a_{35} \end{bmatrix}.$$

What about another vector $\vec{x}$ that is not in the standard basis? Where does it go? We use linearity. First, we write the vector as a linear combination of the standard basis vectors:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3 + x_4\vec{e}_4 + x_5\vec{e}_5.$$

Then

$$A\vec{x} = A(x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3 + x_4\vec{e}_4 + x_5\vec{e}_5) = x_1A\vec{e}_1 + x_2A\vec{e}_2 + x_3A\vec{e}_3 + x_4A\vec{e}_4 + x_5A\vec{e}_5.$$

If we know where A takes all the basis vectors, we know where it takes all vectors.

Suppose M is the 2×2 matrix from above, then

$$M \begin{bmatrix} -2 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 0.1 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 0.1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -1.8 \\ -5.6 \end{bmatrix}.$$

Every linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$ can be represented by an $n \times m$ matrix. You just figure out where it takes the standard basis vectors. Conversely, every $n \times m$ matrix represents a linear mapping. Hence, we may think of matrices being linear mappings, and linear mappings being matrices.

Or can we? In this book we study mostly linear differential operators, and linear differential operators are linear mappings, although they are not acting on $\mathbb{R}^n$, but on an infinite-dimensional space of functions:

$$Lf = g.$$

For a function f we get a function g, and L is linear in the sense that

$$L(f + h) = Lf + Lh, \qquad \text{and} \qquad L(\alpha f) = \alpha Lf.$$

for any number (scalar) α and all functions f and h.

So the answer is not really. But if we consider vectors in finite-dimensional spaces $\mathbb{R}^n$ then yes, every linear mapping is a matrix. We have mentioned at the beginning of this section, that we can "make everything a vector." That's not strictly true, but it is true approximately. Those "infinite-dimensional" spaces of functions can be approximated by a finite-dimensional space, and then linear operators are just matrices. So approximately, this is true. And as far as actual computations that we can do on a computer, we can work only with finitely many dimensions anyway. If you ask a computer or your calculator to plot a function, it samples the function at finitely many points and then connects the dots*. It does not actually give you infinitely many values. The way that you have been using the computer or your calculator so far has already been a certain approximation of the space of functions by a finite-dimensional space.

To end the section, we notice how $A\vec{x}$ can be written more succintly. Suppose

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad \text{and} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then

$$A\vec{x} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) \\ 3 \cdot 2 + 4 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

That is, you take the entries in a row of the matrix, you multiply them by the entries in your vector, you add things up, and that's the corresponding entry in the resulting vector.

A.1.3 Exercises

Exercise* A.1.1*: *On a piece of graph paper draw the vectors:*

a) $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ *b)* $\begin{bmatrix} -2 \\ -4 \end{bmatrix}$ *c)* $(3, -4)$

Exercise* A.1.2*: *On a piece of graph paper draw the vector* $(1, 2)$ *starting at (based at) the given point:*

a) based at $(0, 0)$ *b) based at* $(1, 2)$ *c) based at* $(0, -1)$

Exercise* A.1.3*: *On a piece of graph paper draw the following operations. Draw and label the vectors involved in the operations as well as the result:*

a) $\begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ *b)* $\begin{bmatrix} -3 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ *c)* $3 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Exercise* A.1.4*: *Compute the magnitude of*

a) $\begin{bmatrix} 7 \\ 2 \end{bmatrix}$ *b)* $\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$ *c)* $(1, 3, -4)$

*If you have ever used Matlab, you may have noticed that to plot a function, we take a vector of inputs, ask Matlab to compute the corresponding vector of values of the function, and then we ask it to plot the result.

Exercise* A.1.5*: *Compute*

a) $\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 7 \\ -8 \end{bmatrix}$ *b)* $\begin{bmatrix} -2 \\ 3 \end{bmatrix} - \begin{bmatrix} 6 \\ -4 \end{bmatrix}$ *c)* $-\begin{bmatrix} -3 \\ 2 \end{bmatrix}$

d) $4\begin{bmatrix} -1 \\ 5 \end{bmatrix}$ *e)* $5\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 9\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *f)* $3\begin{bmatrix} 1 \\ -8 \end{bmatrix} - 2\begin{bmatrix} 3 \\ -1 \end{bmatrix}$

Exercise* A.1.6*: *Find the unit vector in the direction of the given vector*

a) $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ *b)* $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$ *c)* $(3, 1, -2)$

Exercise* A.1.7*: *If* $\vec{x} = (1, 2)$ *and* $\vec{y}$ *are added together, we find* $\vec{x} + \vec{y} = (0, 2)$. *What is* $\vec{y}$*?*

Exercise* A.1.8*: *Write* $(1, 2, 3)$ *as a linear combination of the standard basis vectors* $\vec{e}_1$, $\vec{e}_2$, *and* $\vec{e}_3$.

Exercise* A.1.9*: *If the magnitude of* $\vec{x}$ *is* 4*, what is the magnitude of*

a) $0\vec{x}$ *b)* $3\vec{x}$ *c)* $-\vec{x}$ *d)* $-4\vec{x}$ *e)* $\vec{x} + \vec{x}$ *f)* $\vec{x} - \vec{x}$

Exercise* A.1.10*: *Suppose a linear mapping* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *takes* $(1, 0)$ *to* $(2, -1)$ *and it takes* $(0, 1)$ *to* $(3, 3)$. *Where does it take*

a) $(1, 1)$ *b)* $(2, 0)$ *c)* $(2, -1)$

Exercise* A.1.11*: *Suppose a linear mapping* $F : \mathbb{R}^3 \to \mathbb{R}^2$ *takes* $(1, 0, 0)$ *to* $(2, 1)$*, it takes* $(0, 1, 0)$ *to* $(3, 4)$*, and it takes* $(0, 0, 1)$ *to* $(5, 6)$. *Write down the matrix representing the mapping* F.

Exercise* A.1.12*: *Suppose that a mapping* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *takes* $(1, 0)$ *to* $(1, 2)$*,* $(0, 1)$ *to* $(3, 4)$*, and* $(1, 1)$ *to* $(0, -1)$. *Explain why* F *is not linear.*

Exercise* A.1.13** (challenging):*** *Let* $\mathbb{R}^3$ *represent the space of quadratic polynomials in* t*: a point* (a_0, a_1, a_2) *in* $\mathbb{R}^3$ *represents the polynomial* $a_0 + a_1 t + a_2 t^2$. *Consider the derivative* $\frac{d}{dt}$ *as a mapping of* $\mathbb{R}^3$ *to* $\mathbb{R}^3$*, and note that* $\frac{d}{dt}$ *is linear. Write down* $\frac{d}{dt}$ *as a* 3×3 *matrix.*

Exercise* A.1.101*: *Compute the magnitude of*

a) $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ *b)* $\begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$ *c)* $(-2, 1, -2)$

Exercise* A.1.102*: *Find the unit vector in the direction of the given vector*

a) $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ *c)* $(2, -5, 2)$

***Exercise* A.1.103:** *Compute*

a) $\begin{bmatrix}3\\1\end{bmatrix}+\begin{bmatrix}6\\-3\end{bmatrix}$ *b)* $\begin{bmatrix}-1\\2\end{bmatrix}-\begin{bmatrix}2\\-1\end{bmatrix}$ *c)* $-\begin{bmatrix}-5\\3\end{bmatrix}$

d) $2\begin{bmatrix}-2\\4\end{bmatrix}$ *e)* $3\begin{bmatrix}1\\0\end{bmatrix}+7\begin{bmatrix}0\\1\end{bmatrix}$ *f)* $2\begin{bmatrix}2\\-3\end{bmatrix}-6\begin{bmatrix}2\\-1\end{bmatrix}$

***Exercise* A.1.104:** *If the magnitude of $\vec{x}$ is 5, what is the magnitude of*

a) $4\vec{x}$ *b)* $-2\vec{x}$ *c)* $-4\vec{x}$

***Exercise* A.1.105:** *Suppose a linear mapping $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ takes $(1,0)$ to $(1,-1)$ and it takes $(0,1)$ to $(2,0)$. Where does it take*

a) $(1,1)$ *b)* $(0,2)$ *c)* $(1,-1)$

A.2 Matrix algebra

Note: 2–3 lectures

A.2.1 One-by-one matrices

Let us motivate what we want to achieve with matrices. Real-valued linear mappings of the real line, linear functions that eat numbers and spit out numbers, are just multiplications by a number. Consider a mapping defined by multiplying by a number. Let's call this number α. The mapping then takes x to αx. We can *add* such mappings: If we have another mapping β, then

$$\alpha x + \beta x = (\alpha + \beta)x.$$

We get a new mapping $\alpha + \beta$ that multiplies x by, well, $\alpha + \beta$. If D is a mapping that doubles its input, $Dx = 2x$, and T is a mapping that triples, $Tx = 3x$, then $D + T$ is a mapping that multiplies by 5, $(D + T)x = 5x$.

Similarly we can *compose* such mappings, that is, we could apply one and then the other. We take x, we run it through the first mapping α to get α times x, then we run αx through the second mapping β. In other words,

$$\beta(\alpha x) = (\beta\alpha)x.$$

We just multiply those two numbers. Using our doubling and tripling mappings, if we double and then triple, that is $T(Dx)$ then we obtain $3(2x) = 6x$. The composition TD is the mapping that multiplies by 6. For larger matrices, composition also ends up being a kind of multiplication.

A.2.2 Matrix addition and scalar multiplication

The mappings that multiply numbers by numbers are just 1×1 matrices. The number α above could be written as a matrix $[\alpha]$. Perhaps we would want to do to all matrices the same things that we did to those 1×1 matrices at the start of this section above. First, let us add matrices. If we have a matrix A and a matrix B that are of the same size, say $m \times n$, then they are mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. The mapping $A + B$ should also be a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$, and it should do the following to vectors:

$$(A + B)\vec{x} = A\vec{x} + B\vec{x}.$$

It turns out you just add the matrices element-wise: If the ij^{th} entry of A is a_{ij}, and the ij^{th} entry of B is b_{ij}, then the ij^{th} entry of $A + B$ is $a_{ij} + b_{ij}$. If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix},$$

then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix}.$$

Let us illustrate on a more concrete example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & -1 \end{bmatrix} = \begin{bmatrix} 1+7 & 2+8 \\ 3+9 & 4+10 \\ 5+11 & 6-1 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 12 & 14 \\ 16 & 5 \end{bmatrix}.$$

Let's check that this does the right thing to a vector. Let's use some of the vector algebra that we already know, and regroup things:

$$\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} &= \left(2 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \right) + \left(2 \begin{bmatrix} 7 \\ 9 \\ 11 \end{bmatrix} - \begin{bmatrix} 8 \\ 10 \\ -1 \end{bmatrix} \right) \\
&= 2 \left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + \begin{bmatrix} 7 \\ 9 \\ 11 \end{bmatrix} \right) - \left(\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 8 \\ 10 \\ -1 \end{bmatrix} \right) \\
&= 2 \begin{bmatrix} 1+7 \\ 3+9 \\ 5+11 \end{bmatrix} - \begin{bmatrix} 2+8 \\ 4+10 \\ 6-1 \end{bmatrix} = 2 \begin{bmatrix} 8 \\ 12 \\ 16 \end{bmatrix} - \begin{bmatrix} 10 \\ 14 \\ 5 \end{bmatrix} \\
&= \begin{bmatrix} 8 & 10 \\ 12 & 14 \\ 16 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \left(= \begin{bmatrix} 2(8) - 10 \\ 2(12) - 14 \\ 2(16) - 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \\ 27 \end{bmatrix} \right).
\end{aligned}$$

If we replaced the numbers by letters that would constitute a proof! Notice that we did not really have to compute what the result is to convince ourselves that the two expressions were equal.

If the sizes of the matrices do not match, then addition is undefined. If A is 3×2 and B is 2×5, then we cannot add the matrices. We do not know what that could possibly mean.

It is also useful to have a matrix that when added to any other matrix does nothing. This is the zero matrix, the matrix of all zeros:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

We often denote the zero matrix by 0 without specifying size. We would then just write $A + 0$, where we just assume that 0 is the zero matrix of the same size as A.

There are really two things we can multiply matrices by. We can multiply matrices by scalars or we can multiply by other matrices. Let us first consider multiplication by scalars. For a matrix A and a scalar α, we want αA to be the matrix that accomplishes

$$(\alpha A)\vec{x} = \alpha(A\vec{x}).$$

That is just scaling the result by α. If you think about it, scaling every term in A by α achieves just that: If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \qquad \text{then} \qquad \alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \end{bmatrix}.$$

For example,

$$2 \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}.$$

Let us list some properties of matrix addition and scalar multiplication. Denote by 0 the zero matrix, by α, β scalars, and by A, B, C matrices. Then:

$$\begin{aligned} A + 0 = A &= 0 + A, \\ A + B &= B + A, \\ (A + B) + C &= A + (B + C), \\ \alpha(A + B) &= \alpha A + \alpha B, \\ (\alpha + \beta)A &= \alpha A + \beta A. \end{aligned}$$

These rules should look very familiar.

A.2.3 Matrix multiplication

As we mentioned above, composition of linear mappings is also a multiplication of matrices. Suppose A is an $m \times n$ matrix, that is, A takes $\mathbb{R}^n$ to $\mathbb{R}^m$, and B is an $n \times p$ matrix, that is, B takes $\mathbb{R}^p$ to $\mathbb{R}^n$. The composition AB should work as follows

$$AB\vec{x} = A(B\vec{x}).$$

First, a vector $\vec{x}$ in $\mathbb{R}^p$ gets taken to the vector $B\vec{x}$ in $\mathbb{R}^n$. Then the mapping A takes it to the vector $A(B\vec{x})$ in $\mathbb{R}^m$. In other words, the composition AB should be an $m \times p$ matrix. In terms of sizes we should have

$$\text{“} \quad [m \times n]\,[n \times p] = [m \times p]. \quad \text{”}$$

Notice how the middle size must match.

OK, now we know what sizes of matrices we should be able to multiply, and what the product should be. Let us see how to actually compute matrix multiplication. We start with the so-called *dot product* (or *inner product*) of two vectors. Usually this is a row vector multiplied with a column vector of the same size. Dot product multiplies each pair of entries from the first and the second vector and sums these products. The result is a single number. For example,

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

And similarly for larger (or smaller) vectors. A dot product is really a product of two matrices: a $1 \times n$ matrix and an $n \times 1$ matrix resulting in a 1×1 matrix, that is, a number.

Armed with the dot product we define the *product of matrices*. We denote by $\operatorname{row}_i(A)$ the i^{th} row of A and by $\operatorname{column}_j(A)$ the j^{th} column of A. For an $m \times n$ matrix A and an $n \times p$ matrix B we can compute the product AB: The matrix AB is an $m \times p$ matrix whose ij^{th} entry is the dot product

$$\operatorname{row}_i(A) \cdot \operatorname{column}_j(B).$$

For example, given a 2×3 and a 3×2 matrix we should end up with a 2×2 matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}, \tag{A.1}$$

or with some numbers:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -7 & 0 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-1) + 2 \cdot (-7) + 3 \cdot 1 & 1 \cdot 2 + 2 \cdot 0 + 3 \cdot (-1) \\ 4 \cdot (-1) + 5 \cdot (-7) + 6 \cdot 1 & 4 \cdot 2 + 5 \cdot 0 + 6 \cdot (-1) \end{bmatrix} = \begin{bmatrix} -12 & -1 \\ -33 & 2 \end{bmatrix}.$$

A useful consequence of the definition is that the evaluation $A\vec{x}$ for a matrix A and a (column) vector $\vec{x}$ is also matrix multiplication. That is really why we think of vectors as column vectors, or $n \times 1$ matrices. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) \\ 3 \cdot 2 + 4 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

If you look at the last section, that is precisely the last example we gave.

You should stare at the computation of multiplication of matrices AB and the previous definition of $A\vec{y}$ as a mapping for a moment. What we are doing with matrix multiplication is applying the mapping A to the columns of B. This is usually written as follows. Suppose we write the $n \times p$ matrix $B = [\vec{b}_1\ \vec{b}_2\ \cdots\ \vec{b}_p]$, where $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p$ are the columns of B. Then for an $m \times n$ matrix A,

$$AB = A[\vec{b}_1\ \vec{b}_2\ \cdots\ \vec{b}_p] = [A\vec{b}_1\ A\vec{b}_2\ \cdots\ A\vec{b}_p].$$

The columns of the $m \times p$ matrix AB are the vectors $A\vec{b}_1, A\vec{b}_2, \ldots, A\vec{b}_p$. For example, in (A.1), the columns of

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

are

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix}.$$

This is a very useful way to understand what matrix multiplication is. It should also make it easier to remember how to perform matrix multiplication.

A.2.4 Some rules of matrix algebra

For multiplication we want an analogue of a 1. That is, we desire a matrix that just leaves everything as it found it. This analogue is the so-called *identity matrix*. The identity matrix is a square matrix with 1s on the main diagonal and zeros everywhere else. It is usually denoted by I. For each size we have a different identity matrix and so sometimes we may denote the size as a subscript. For example, I_3 is the 3×3 identity matrix

$$I = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us see how the matrix works on a smaller example,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} \cdot 1 + a_{12} \cdot 0 & a_{11} \cdot 0 + a_{12} \cdot 1 \\ a_{21} \cdot 1 + a_{22} \cdot 0 & a_{21} \cdot 0 + a_{22} \cdot 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Multiplication by the identity from the left looks similar, and also does not touch anything.

We have the following rules for matrix multiplication. Suppose that A, B, C are matrices of the correct sizes so that the following make sense. Let α denote a scalar (number). Then

$$\begin{aligned} A(BC) &= (AB)C && \text{(associative law)}, \\ A(B+C) &= AB + AC && \text{(distributive law)}, \\ (B+C)A &= BA + CA && \text{(distributive law)}, \\ \alpha(AB) &= (\alpha A)B = A(\alpha B), && \\ IA &= A = AI && \text{(identity)}. \end{aligned}$$

Example A.2.1: Let us demonstrate a couple of these rules. For example, the associative law:

$$\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \Bigg(\underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} \Bigg) = \underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 16 & 24 \\ -16 & -2 \end{bmatrix}}_{BC} = \underbrace{\begin{bmatrix} -96 & -78 \\ 64 & 52 \end{bmatrix}}_{A(BC)},$$

and

$$\Bigg(\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \Bigg) \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} = \underbrace{\begin{bmatrix} -9 & -21 \\ 6 & 14 \end{bmatrix}}_{AB} \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} = \underbrace{\begin{bmatrix} -96 & -78 \\ 64 & 52 \end{bmatrix}}_{(AB)C}.$$

Or how about multiplication by scalars:

$$10 \Bigg(\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \Bigg) = 10 \underbrace{\begin{bmatrix} -9 & -21 \\ 6 & 14 \end{bmatrix}}_{AB} = \underbrace{\begin{bmatrix} -90 & -210 \\ 60 & 140 \end{bmatrix}}_{10(AB)},$$

$$\underbrace{\left(10\begin{bmatrix}-3 & 3\\ 2 & -2\end{bmatrix}\right)}_{A}\underbrace{\begin{bmatrix}4 & 4\\ 1 & -3\end{bmatrix}}_{B} = \underbrace{\begin{bmatrix}-30 & 30\\ 20 & -20\end{bmatrix}}_{10A}\underbrace{\begin{bmatrix}4 & 4\\ 1 & -3\end{bmatrix}}_{B} = \underbrace{\begin{bmatrix}-90 & -210\\ 60 & 140\end{bmatrix}}_{(10A)B},$$

and

$$\underbrace{\begin{bmatrix}-3 & 3\\ 2 & -2\end{bmatrix}}_{A}\left(10\underbrace{\begin{bmatrix}4 & 4\\ 1 & -3\end{bmatrix}}_{B}\right) = \underbrace{\begin{bmatrix}-3 & 3\\ 2 & -2\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}40 & 40\\ 10 & -30\end{bmatrix}}_{10B} = \underbrace{\begin{bmatrix}-90 & -210\\ 60 & 140\end{bmatrix}}_{A(10B)}.$$

A multiplication rule, one you have used since primary school on numbers, is quite conspicuously missing for matrices. That is, matrix multiplication is not commutative. Firstly, just because AB makes sense, it may be that BA is not even defined. For example, if A is 2×3, and B is 3×4, the we can multiply AB but not BA.

Even if AB and BA are both defined, does not mean that they are equal. For example, take $A = \left[\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}1 & 0\\ 0 & 2\end{smallmatrix}\right]$:

$$AB = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix} = \begin{bmatrix}1 & 2\\ 1 & 2\end{bmatrix} \qquad \neq \qquad \begin{bmatrix}1 & 1\\ 2 & 2\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} = BA.$$

A.2.5 Inverse

A couple of other algebra rules you know for numbers do not quite work on matrices:

(i) $AB = AC$ does not necessarily imply $B = C$, even if A is not 0.

(ii) $AB = 0$ does not necessarily mean that $A = 0$ or $B = 0$.

For example:

$$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}0 & 2\\ 0 & 0\end{bmatrix}.$$

To make these rules hold, we do not just need one of the matrices to not be zero, we would need to "divide" by a matrix. This is where the *matrix inverse* comes in. Suppose that A and B are $n \times n$ matrices such that

$$AB = I = BA.$$

Then we call B the inverse of A and we denote B by A^{-1}. Perhaps not surprisingly, $(A^{-1})^{-1} = A$, since if the inverse of A is B, then the inverse of B is A. If the inverse of A exists, then we say A is *invertible*. If A is not invertible, we say A is *singular*.

If $A = [a]$ is a 1×1 matrix, then A^{-1} is $a^{-1} = \frac{1}{a}$. That is where the notation comes from. The computation is not nearly as simple when A is larger.

The proper formulation of the cancellation rule is:

If A is invertible, then $AB = AC$ implies $B = C$.

The computation is what you would do in regular algebra with numbers, but you have to be careful never to commute matrices:

$$\begin{aligned} AB &= AC, \\ A^{-1}AB &= A^{-1}AC, \\ IB &= IC, \\ B &= C. \end{aligned}$$

And similarly for cancellation on the right:

If A is invertible, then $BA = CA$ implies $B = C$.

The rule says, among other things, that the inverse of a matrix is unique if it exists: If $AB = I = AC$, then A is invertible and $B = C$.

We will see later how to compute an inverse of a matrix in general. For now, let us note that there is a simple formula for the inverse of a 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

For example:

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}^{-1} = \frac{1}{1 \cdot 4 - 1 \cdot 2} \begin{bmatrix} 4 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix}.$$

Let's try it:

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Just as we cannot divide by every number, not every matrix is invertible. In the case of matrices however we may have singular matrices that are not zero. For example,

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

is a singular matrix. But didn't we just give a formula for an inverse? Let us try it:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}^{-1} = \frac{1}{1 \cdot 2 - 1 \cdot 2} \begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} = ?$$

We get into a bit of trouble; we are trying to divide by zero.

So a 2×2 matrix A is invertible whenever

$$ad - bc \neq 0$$

and otherwise it is singular. The expression $ad - bc$ is called the *determinant* and we will look at it more carefully in a later section. There is a similar expression for a square matrix of any size.

A.2.6 Diagonal matrices

A simple (and surprisingly useful) type of a square matrix is a so-called *diagonal matrix*. It is a matrix whose entries are all zero except those on the main diagonal from top left to bottom right. For example a 4×4 diagonal matrix is of the form

$$\begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}.$$

Such matrices have nice properties when we multiply by them. If we multiply them by a vector, they multiply the k^{th} entry by d_k. For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 \\ 2 \cdot 5 \\ 3 \cdot 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \\ 18 \end{bmatrix}.$$

Similarly, when they multiply another matrix from the left, they multiply the k^{th} row by d_k. For example,

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

On the other hand, multiplying on the right, they multiply the columns:

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

And it is really easy to multiply two diagonal matrices together—we multiply the entries:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 & 0 & 0 \\ 0 & 2 \cdot 3 & 0 \\ 0 & 0 & 3 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

For this last reason, they are easy to invert, you simply invert each diagonal element:

$$\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}^{-1} = \begin{bmatrix} d_1^{-1} & 0 & 0 \\ 0 & d_2^{-1} & 0 \\ 0 & 0 & d_3^{-1} \end{bmatrix}.$$

Let us check an example

$$\underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}^{-1}}_{A^{-1}} \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{I}.$$

It is no wonder that the way we solve many problems in linear algebra (and in differential equations) is to try to reduce the problem to the case of diagonal matrices.

A.2.7 Transpose

Vectors do not always have to be column vectors, that is just a convention. Swapping rows and columns is from time to time needed. The operation that swaps rows and columns is the so-called *transpose*. The transpose of A is denoted by A^T. Example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Transpose takes an $m \times n$ matrix to an $n \times m$ matrix.

A key feature of the transpose is that if the product AB makes sense, then $B^T A^T$ also makes sense, at least from the point of view of sizes. In fact, we get precisely the transpose of AB. That is:

$$(AB)^T = B^T A^T.$$

For example,

$$\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & -2 \end{bmatrix} \right)^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

It is left to the reader to verify that computing the matrix product on the left and then transposing is the same as computing the matrix product on the right.

If we have a column vector $\vec{x}$ to which we apply a matrix A and we transpose the result, then the row vector $\vec{x}^T$ applies to A^T from the left:

$$(A\vec{x})^T = \vec{x}^T A^T.$$

Another place where transpose is useful is when we wish to apply the dot product* to two column vectors:

$$\vec{x} \cdot \vec{y} = \vec{y}^T \vec{x}.$$

That is the way that one often writes the dot product in software.

We say a matrix A is *symmetric* if $A = A^T$. For example,

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

is a symmetric matrix. Notice that a symmetric matrix is always square, that is, $n \times n$. Symmetric matrices have many nice properties†, and come up quite often in applications.

*As a side note, mathematicians write $\vec{y}^T \vec{x}$ and physicists write $\vec{x}^T \vec{y}$. Shhh. . . don't tell anyone, but the physicists are probably right on this.

†Although so far we have not learned enough about matrices to really appreciate them.

A.2.8 Exercises

***Exercise* A.2.1:** *Add the following matrices*

a) $\begin{bmatrix} -1 & 2 & 2 \\ 5 & 8 & -1 \end{bmatrix} + \begin{bmatrix} 3 & 2 & 3 \\ 8 & 3 & 5 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 0 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -8 & -3 \\ 3 & 1 & 0 \\ 6 & -4 & 1 \end{bmatrix}$

***Exercise* A.2.2:** *Compute*

a) $3 \begin{bmatrix} 0 & 3 \\ -2 & 2 \end{bmatrix} + 6 \begin{bmatrix} 1 & 5 \\ -1 & 5 \end{bmatrix}$ *b)* $2 \begin{bmatrix} -3 & 1 \\ 2 & 2 \end{bmatrix} - 3 \begin{bmatrix} 2 & -1 \\ 3 & 2 \end{bmatrix}$

***Exercise* A.2.3:** *Multiply the following matrices*

a) $\begin{bmatrix} -1 & 2 \\ 3 & 1 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} 3 & -1 & 3 & 1 \\ 8 & 3 & 2 & -3 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 & 7 \\ 1 & 2 & 3 & -1 \\ 1 & -1 & 3 & 0 \end{bmatrix}$

c) $\begin{bmatrix} 4 & 1 & 6 & 3 \\ 5 & 6 & 5 & 0 \\ 4 & 6 & 6 & 0 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 2 \\ 3 & 5 \\ 5 & 6 \end{bmatrix}$ *d)* $\begin{bmatrix} 1 & 1 & 4 \\ 0 & 5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & 0 \\ 6 & 4 \end{bmatrix}$

***Exercise* A.2.4:** *Compute the inverse of the given matrices*

a) $\begin{bmatrix} -3 \end{bmatrix}$ *b)* $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 4 \\ 1 & 3 \end{bmatrix}$ *d)* $\begin{bmatrix} 2 & 2 \\ 1 & 4 \end{bmatrix}$

***Exercise* A.2.5:** *Compute the inverse of the given matrices*

a) $\begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ *b)* $\begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}$

***Exercise* A.2.101:** *Add the following matrices*

a) $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 5 & 3 & 4 \\ 1 & 2 & 5 \end{bmatrix}$ *b)* $\begin{bmatrix} 6 & -2 & 3 \\ 7 & 3 & 3 \\ 8 & -1 & 2 \end{bmatrix} + \begin{bmatrix} -1 & -1 & -3 \\ 6 & 7 & 3 \\ -9 & 4 & -1 \end{bmatrix}$

***Exercise* A.2.102:** *Compute*

a) $2 \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + 3 \begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix}$ *b)* $3 \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} - 2 \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$

Exercise* A.2.103*: *Multiply the following matrices*

a) $\begin{bmatrix} 2 & 1 & 4 \\ 3 & 4 & 4 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 6 & 3 \\ 3 & 5 \end{bmatrix}$

b) $\begin{bmatrix} 0 & 3 & 3 \\ 2 & -2 & 1 \\ 3 & 5 & -2 \end{bmatrix} \begin{bmatrix} 6 & 6 & 2 \\ 4 & 6 & 0 \\ 2 & 0 & 4 \end{bmatrix}$

c) $\begin{bmatrix} 3 & 4 & 1 \\ 2 & -1 & 0 \\ 4 & -1 & 5 \end{bmatrix} \begin{bmatrix} 0 & 2 & 5 & 0 \\ 2 & 0 & 5 & 2 \\ 3 & 6 & 1 & 6 \end{bmatrix}$

d) $\begin{bmatrix} -2 & -2 \\ 5 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & 3 \end{bmatrix}$

Exercise* A.2.104*: *Compute the inverse of the given matrices*

a) $\begin{bmatrix} 2 \end{bmatrix}$ *b)* $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ *d)* $\begin{bmatrix} 4 & 2 \\ 4 & 4 \end{bmatrix}$

Exercise* A.2.105*: *Compute the inverse of the given matrices*

a) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ *b)* $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ *c)* $\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$

A.3 Elimination

Note: 2–3 lectures

A.3.1 Linear systems of equations

One application of matrices is to solve systems of linear equations*. Consider the following system of linear equations

$$\begin{aligned} 2x_1 + 2x_2 + 2x_3 &= 2, \\ x_1 + x_2 + 3x_3 &= 5, \\ x_1 + 4x_2 + x_3 &= 10. \end{aligned} \tag{A.2}$$

There is a systematic procedure called *elimination* to solve such a system. In this procedure, we attempt to eliminate each variable from all but one equation. We want to end up with equations such as $x_3 = 2$, where we can just read off the answer.

We write a system of linear equations as a matrix equation:

$$A\vec{x} = \vec{b}.$$

The system (A.2) is written as

$$\underbrace{\begin{bmatrix} 2 & 2 & 2 \\ 1 & 1 & 3 \\ 1 & 4 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix}}_{\vec{b}} .$$

If we knew the inverse of A, then we would be done; we would simply solve the equation:

$$\vec{x} = A^{-1}A\vec{x} = A^{-1}\vec{b}.$$

Well, but that is part of the problem, we do not know how to compute the inverse for matrices bigger than 2×2. We will see later that to compute the inverse we are really solving $A\vec{x} = \vec{b}$ for several different $\vec{b}$. In other words, we will need to do elimination to find A^{-1}. In addition, we may wish to solve $A\vec{x} = \vec{b}$ if A is not invertible, or perhaps not even square.

Let us return to the equations themselves and see how we can manipulate them. There are a few operations we can perform on the equations that do not change the solution. First, perhaps an operation that may seem stupid, we can swap two equations in (A.2):

$$\begin{aligned} x_1 + x_2 + 3x_3 &= 5, \\ 2x_1 + 2x_2 + 2x_3 &= 2, \\ x_1 + 4x_2 + x_3 &= 10. \end{aligned}$$

*Although perhaps we have this backwards, quite often we solve a linear system of equations to find out something about matrices, rather than vice versa.

Clearly these new equations have the same solutions x_1, x_2, x_3. A second operation is that we can multiply an equation by a nonzero number. For example, we multiply the third equation in (A.2) by 3:

$$\begin{aligned} 2x_1 + \ \ 2x_2 + 2x_3 &= 2, \\ x_1 + \ \ \ x_2 + 3x_3 &= 5, \\ 3x_1 + 12x_2 + 3x_3 &= 30. \end{aligned}$$

Finally, we can add a multiple of one equation to another equation. For instance, we add 3 times the third equation in (A.2) to the second equation:

$$\begin{aligned} 2x_1 + \quad\quad\quad 2x_2 + \quad\quad\quad 2x_3 &= 2, \\ (1+3)x_1 + \ (1+12)x_2 + \ (3+3)x_3 &= 5+30, \\ x_1 + \quad\quad\quad 4x_2 + \quad\quad\quad x_3 &= 10. \end{aligned}$$

The same x_1, x_2, x_3 should still be solutions to the new equations. These were just examples; we did not get any closer to the solution. We must to do these three operations in some more logical manner, but it turns out these three operations suffice to solve every linear equation.

The first thing is to write the equations in a more compact manner. Given

$$A\vec{x} = \vec{b},$$

we write down the so-called *augmented matrix*

$$[A \mid \vec{b}],$$

where the vertical line is just a marker for us to know where the "right-hand side" of the equation starts. For the system (A.2) the augmented matrix is

$$\left[\begin{array}{ccc|c} 2 & 2 & 2 & 2 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{array}\right].$$

The entire process of elimination, which we will describe, is often applied to any sort of matrix, not just an augmented matrix. Simply think of the matrix as the 3×4 matrix

$$\begin{bmatrix} 2 & 2 & 2 & 2 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{bmatrix}.$$

A.3.2 Row echelon form and elementary operations

We apply the three operations above to the matrix. We call these the *elementary operations* or *elementary row operations*. Translating the operations to the matrix setting, the operations become:

(i) Swap two rows.

(ii) Multiply a row by a nonzero number.

(iii) Add a multiple of one row to another row.

We run these operations until we get into a state where it is easy to read off the answer, or until we get into a contradiction indicating no solution.

More specifically, we run the operations until we obtain the so-called *row echelon form*. Let us call the first (from the left) nonzero entry in each row the *leading entry*. A matrix is in *row echelon form* if the following conditions are satisfied:

(i) The leading entry in any row is strictly to the right of the leading entry of the row above.

(ii) Any zero rows are below all the nonzero rows.

(iii) All leading entries are 1.

A matrix is in *reduced row echelon form* if furthermore the following condition is satisfied.

(iv) All the entries above a leading entry are zero.

Note that the definition applies to matrices of any size.

Example A.3.1: The following matrices are in row echelon form. The leading entries are marked:

$$\begin{bmatrix} \boxed{1} & 2 & 9 & 3 \\ 0 & 0 & \boxed{1} & 5 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix} \qquad \begin{bmatrix} \boxed{1} & -1 & -3 \\ 0 & \boxed{1} & 5 \\ 0 & 0 & \boxed{1} \end{bmatrix} \qquad \begin{bmatrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{1} & 2 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & \boxed{1} & -5 & 2 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

None of the matrices above are in *reduced* row echelon form. For example, in the first matrix none of the entries above the second and third leading entries are zero; they are 9, 3, and 5. The following matrices are in reduced row echelon form. The leading entries are marked:

$$\begin{bmatrix} \boxed{1} & 3 & 0 & 8 \\ 0 & 0 & \boxed{1} & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} \boxed{1} & 0 & 2 & 0 \\ 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix} \qquad \begin{bmatrix} \boxed{1} & 0 & 3 \\ 0 & \boxed{1} & -2 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & \boxed{1} & 2 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The procedure we will describe to find a reduced row echelon form of a matrix is called *Gauss–Jordan elimination*. The first part of it, which obtains a row echelon form, is called *Gaussian elimination* or *row reduction*. For some problems, a row echelon form is sufficient, and it is a bit less work to only do this first part.

To attain the row echelon form we work systematically. We go column by column, starting at the first column. We find topmost entry in the first column that is not zero, and we call it the *pivot*. If there is no nonzero entry we move to the next column. We swap rows to put the row with the pivot as the first row. We divide the first row by the pivot to make the pivot entry be a 1. Now look at all the rows below and subtract the correct multiple of the pivot row so that all the entries below the pivot become zero.

After this procedure we forget that we had a first row (it is now fixed), and we forget about the column with the pivot and all the preceding zero columns. Below the pivot row, all the entries in these columns are just zero. Then we focus on the smaller matrix and we repeat the steps above.

It is best shown by example, so let us go back to the example from the beginning of the section. We keep the vertical line in the matrix, even though the procedure works on any matrix, not just an augmented matrix. We start with the first column and we locate the pivot, in this case the first entry of the first column.

$$\left[\begin{array}{ccc|c} \boxed{2} & 2 & 2 & 2 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{array}\right]$$

We multiply the first row by $1/2$.

$$\left[\begin{array}{ccc|c} \boxed{1} & 1 & 1 & 1 \\ 1 & 1 & 3 & 5 \\ 1 & 4 & 1 & 10 \end{array}\right]$$

We subtract the first row from the second and third row (two elementary operations).

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & 3 & 0 & 9 \end{array}\right]$$

We are done with the first column and the first row for now. We almost pretend the matrix does not have the first column and the first row.

$$\left[\begin{array}{ccc|c} * & * & * & * \\ * & 0 & 2 & 4 \\ * & 3 & 0 & 9 \end{array}\right]$$

OK, look at the second column, and notice that now the pivot is in the third row.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & \boxed{3} & 0 & 9 \end{array}\right]$$

We swap rows.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & \boxed{3} & 0 & 9 \\ 0 & 0 & 2 & 4 \end{array}\right]$$

And we divide the pivot row by 3.

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

We do not need to subtract anything as everything below the pivot is already zero. We move on, we again start ignoring the second row and second column and focus on

$$\left[\begin{array}{ccc|c} * & * & * & * \\ * & * & * & * \\ * & * & 2 & 4 \end{array}\right].$$

We find the pivot, then divide that row by 2:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & \boxed{2} & 4 \end{array}\right] \qquad \to \qquad \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

The matrix is now in row echelon form.

The equation corresponding to the last row is $x_3 = 2$. We know x_3 and we could substitute it into the first two equations to get equations for x_1 and x_2. Then we could do the same thing with x_2, until we solve for all 3 variables. This procedure is called *backsubstitution* and we can achieve it via elementary operations. We start from the lowest pivot (leading entry in the row echelon form) and subtract the right multiple from the row above to make all the entries above this pivot zero. Then we move to the next pivot and so on. After we are done, we will have a matrix in reduced row echelon form.

We continue our example. Subtract the last row from the first to get

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

The entry above the pivot in the second row is already zero. So we move onto the next pivot, the one in the second row. We subtract this row from the top row to get

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

The matrix is in reduced row echelon form.

If we now write down the equations for x_1, x_2, x_3, we find

$$x_1 = -4, \qquad x_2 = 3, \qquad x_3 = 2.$$

In other words, we have solved the system.

A.3.3 Non-unique solutions and inconsistent systems

It is possible that the solution of a linear system of equations is not unique, or that no solution exists. Suppose for a moment that the row echelon form we found was

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

Then we have an equation $0 = 1$ coming from the last row. That is impossible and the equations are what we call *inconsistent*. There is no solution to $A\vec{x} = \vec{b}$.

On the other hand, if we find a row echelon form

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

then there is no issue with finding solutions. In fact, we will find way too many. Let us continue with backsubstitution (subtracting 3 times the third row from the first) to find the reduced row echelon form and let's mark the pivots.

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

The last row is all zeros; it just says $0 = 0$ and we ignore it. The two remaining equations are

$$x_1 + 2x_2 = -5, \qquad x_3 = 3.$$

Let us solve for the variables that corresponded to the pivots, that is x_1 and x_3 as there was a pivot in the first column and in the third column:

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

The variable x_2 can be anything you wish and we still get a solution. The x_2 is called a *free variable*. There are infinitely many solutions, one for every choice of x_2. If we pick $x_2 = 0$, then $x_1 = -5$, and $x_3 = 3$ give a solution. But we also get a solution by picking say $x_2 = 1$, in which case $x_1 = -9$ and $x_3 = 3$, or by picking $x_2 = -5$ in which case $x_1 = 5$ and $x_3 = 3$.

The general idea is that if any row has all zeros in the columns corresponding to the variables, but a nonzero entry in the column corresponding to the right-hand side $\vec{b}$, then the system is inconsistent and has no solutions. In other words, the system is inconsistent if you find a pivot on the right side of the vertical line drawn in the augmented matrix. Otherwise, the system is *consistent*, and at least one solution exists.

Suppose the system is consistent (at least one solution exists):

(i) If every column corresponding to a variable has a pivot element, then the solution is unique.

(ii) If there are columns corresponding to variables with no pivot, then those are *free variables* that can be chosen arbitrarily, and there are infinitely many solutions.

When $\vec{b} = \vec{0}$, we have a so-called *homogeneous matrix equation*

$$A\vec{x} = \vec{0}.$$

There is no need to write an augmented matrix in this case. As the elementary operations do not do anything to a zero column, it always stays a zero column. Moreover, $A\vec{x} = \vec{0}$ always has at least one solution, namely $\vec{x} = \vec{0}$. Such a system is always consistent. It may have other solutions: If you find any free variables, then you get infinitely many solutions.

The set of solutions of $A\vec{x} = \vec{0}$ comes up quite often so people give it a name. It is called the *nullspace* or the *kernel* of A. One place where the kernel comes up is invertibility of a square matrix A. If the kernel of A contains a nonzero vector, then it contains infinitely many vectors (there was a free variable). But then it is impossible to invert $\vec{0}$, since infinitely many vectors go to $\vec{0}$, so there is no unique vector that A takes to $\vec{0}$. So if the kernel is nontrivial, that is, if there are any nonzero vectors in the kernel, in other words, if there are any free variables, or in yet other words, if the row echelon form of A has columns without pivots, then A is not invertible. We will return to this idea later.

A.3.4 Linear independence and rank

If rows of a matrix correspond to equations, it may be good to find out how many equations we really need to find the same set of solutions. Similarly, if we find a number of solutions to a linear equation $A\vec{x} = \vec{0}$, we may ask if we found enough so that all other solutions can be formed out of the given set. The concept we want is that of linear independence. That same concept is useful for differential equations, for example in chapter **??**.

Given row or column vectors $\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n$, a *linear combination* is an expression of the form

$$\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2 + \cdots + \alpha_n\vec{y}_n,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are all scalars. For example, $3\vec{y}_1 + \vec{y}_2 - 5\vec{y}_3$ is a linear combination of $\vec{y}_1$, $\vec{y}_2$, and $\vec{y}_3$.

We have seen linear combinations before. The expression

$$A\vec{x}$$

is a linear combination of the columns of A, while

$$\vec{x}^T A = (A^T\vec{x})^T$$

is a linear combination of the rows of A.

The way linear combinations come up in our study of differential equations is similar to the following computation. Suppose that $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are solutions to $A\vec{x}_1 = \vec{0}$, $A\vec{x}_2 = \vec{0}$, $\ldots$, $A\vec{x}_n = \vec{0}$. Then the linear combination

$$\vec{y} = \alpha_1\vec{x}_1 + \alpha_2\vec{x}_2 + \cdots + \alpha_n\vec{x}_n$$

is a solution to $A\vec{y} = \vec{0}$:

$$\begin{aligned} A\vec{y} = A(\alpha_1\vec{x}_1 + \alpha_2\vec{x}_2 + \cdots + \alpha_n\vec{x}_n) = \\ = \alpha_1 A\vec{x}_1 + \alpha_2 A\vec{x}_2 + \cdots + \alpha_n A\vec{x}_n = \alpha_1\vec{0} + \alpha_2\vec{0} + \cdots + \alpha_n\vec{0} = \vec{0}. \end{aligned}$$

So if you have found enough solutions, you have them all. The question is, when did we find enough of them?

We say the vectors $\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n$ are *linearly independent* if the only solution to

$$\alpha_1\vec{x}_1 + \alpha_2\vec{x}_2 + \cdots + \alpha_n\vec{x}_n = \vec{0}$$

is $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Otherwise, we say the vectors are *linearly dependent*.

For example, the vectors $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ are linearly independent. Let's try:

$$\alpha_1 \begin{bmatrix}1\\2\end{bmatrix} + \alpha_2 \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\alpha_1\\2\alpha_1+\alpha_2\end{bmatrix} = \vec{0} = \begin{bmatrix}0\\0\end{bmatrix}.$$

So $\alpha_1 = 0$, and then it is clear that $\alpha_2 = 0$ as well. In other words, the two vectors are linearly independent.

If a set of vectors is linearly dependent, that is, some of the α_js are nonzero, then we can solve for one vector in terms of the others. Suppose $\alpha_1 \neq 0$. Since $\alpha_1\vec{x}_1 + \alpha_2\vec{x}_2 + \cdots + \alpha_n\vec{x}_n = \vec{0}$, then

$$\vec{x}_1 = \frac{-\alpha_2}{\alpha_1}\vec{x}_2 - \frac{-\alpha_3}{\alpha_1}\vec{x}_3 + \cdots + \frac{-\alpha_n}{\alpha_1}\vec{x}_n.$$

For example,

$$2\begin{bmatrix}1\\2\\3\end{bmatrix} - 4\begin{bmatrix}1\\1\\1\end{bmatrix} + 2\begin{bmatrix}1\\0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix},$$

and so

$$\begin{bmatrix}1\\2\\3\end{bmatrix} = 2\begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}1\\0\\-1\end{bmatrix}.$$

You may have noticed that solving for those α_js is just solving linear equations, and so you may not be surprised that to check if a set of vectors is linearly independent we use row reduction.

Given a set of vectors, we may not be interested in just finding if they are linearly independent or not, we may be interested in finding a linearly independent subset. Or perhaps we may want to find some other vectors that give the same linear combinations and are linearly independent. The way to figure this out is to form a matrix out of our vectors. If we have row vectors we consider them as rows of a matrix. If we have column vectors we consider them columns of a matrix. The set of all linear combinations of a set of vectors is called their *span*.

$$\operatorname{span}\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\} = \{\text{Set of all linear combinations of } \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}.$$

Given a matrix A, the maximal number of linearly independent rows is called the *rank* of A, and we write "rank A" for the rank. For example,

$$\operatorname{rank}\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} = 1.$$

The second and third row are multiples of the first one. We cannot choose more than one row and still have a linearly independent set. But what is

$$\operatorname{rank}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad = \quad ?$$

That seems to be a tougher question to answer. The first two rows are linearly independent (neither is a multiple of the other), so the rank is at least two. If we would set up the equations for the α_1, α_2, and α_3, we would find a system with infinitely many solutions. One solution is

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

So the set of all three rows is linearly dependent, the rank cannot be 3. Therefore the rank is 2.

But how can we do this in a more systematic way? We find the row echelon form!

$$\text{Row echelon form of} \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

The elementary row operations do not change the set of linear combinations of the rows (that was one of the main reasons for defining them as they were). In other words, the span of the rows of the A is the same as the span of the rows of the row echelon form of A. In particular, the number of linearly independent rows is the same. And in the row echelon form, all nonzero rows are linearly independent. This is not hard to see. Consider the two nonzero rows in the example above. Suppose we tried to solve for the α_1 and α_2 in

$$\alpha_1 \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

Since the first column of the row echelon matrix has zeros except in the first row means that $\alpha_1 = 0$. For the same reason, α_2 is zero. We only have two nonzero rows, and they are linearly independent, so the rank of the matrix is 2.

The span of the rows is called the *row space*. The row space of A and the row echelon form of A are the same. In the example,

$$\text{row space of } \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \operatorname{span} \left\{ \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}, \begin{bmatrix} 7 & 8 & 9 \end{bmatrix} \right\}$$
$$= \operatorname{span} \left\{ \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \right\}.$$

Similarly to row space, the span of columns is called the *column space*.

$$\text{column space of } \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} \right\}.$$

So it may also be good to find the number of linearly independent columns of A. One way to do that is to find the number of linearly independent rows of A^T. It is a tremendously useful fact that the number of linearly independent columns is always the same as the number of linearly independent rows:

Theorem A.3.1. $\operatorname{rank} A = \operatorname{rank} A^T$

In particular, to find a set of linearly independent columns we need to look at where the pivots were. If you recall above, when solving $A\vec{x} = \vec{0}$ the key was finding the pivots, any non-pivot columns corresponded to free variables. That means we can solve for the non-pivot columns in terms of the pivot columns. Let's see an example. First we reduce some random matrix:

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

We find a pivot and reduce the rows below:

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

We find the next pivot, make it one, and rinse and repeat:

$$\begin{bmatrix} \boxed{1} & 2 & 3 & 4 \\ 0 & 0 & \boxed{-1} & -2 \\ 0 & 0 & -2 & -4 \end{bmatrix} \to \begin{bmatrix} \boxed{1} & 2 & 3 & 4 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & -2 & -4 \end{bmatrix} \to \begin{bmatrix} \boxed{1} & 2 & 3 & 4 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The final matrix is the row echelon form of the matrix. Consider the pivots that we marked. The pivot columns are the first and the third column. All other columns correspond to free variables when solving $A\vec{x} = \vec{0}$, so all other columns can be solved in terms of the first and the third column. In other words

$$\text{column space of } \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 6 \\ 3 & 6 & 7 & 8 \end{bmatrix} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} \right\} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} \right\}.$$

We could perhaps use another pair of columns to get the same span, but the first and the third are guaranteed to work because they are pivot columns.

The discussion above could be expanded into a proof of the theorem if we wanted. As each nonzero row in the row echelon form contains a pivot, then the rank is the number of pivots, which is the same as the maximal number of linearly independent columns.

The idea also works in reverse. Suppose we have a bunch of column vectors and we just need to find a linearly independent set. For example, suppose we started with the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \quad \vec{v}_4 = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}.$$

These vectors are not linearly independent as we saw above. In particular, the span of $\vec{v}_1$ and $\vec{v}_3$ is the same as the span of all four of the vectors. So $\vec{v}_2$ and $\vec{v}_4$ can both be written as linear combinations of $\vec{v}_1$ and $\vec{v}_3$. A common thing that comes up in practice is that one gets a set of vectors whose span is the set of solutions of some problem. But perhaps we get way too many vectors, we want to simplify. For example above, all vectors in the span of $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ can be written $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3 + \alpha_4\vec{v}_4$ for some numbers $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. But it is also true that every such vector can be written as $a\vec{v}_1 + b\vec{v}_3$ for two numbers a and b. And one has to admit, that looks much simpler. Moreover, these numbers a and b are unique. More on that in the next section.

To find this linearly independent set we simply take our vectors and form the matrix $[\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4]$, that is, the matrix

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

We crank up the row-reduction machine, feed this matrix into it, find the pivot columns, and pick those. In this case, $\vec{v}_1$ and $\vec{v}_3$.

A.3.5 Computing the inverse

If the matrix A is square and there exists a unique solution $\vec{x}$ to $A\vec{x} = \vec{b}$ for any $\vec{b}$ (there are no free variables), then A is invertible. This is equivalent to the $n \times n$ matrix A being of rank n.

In particular, if $A\vec{x} = \vec{b}$ then $\vec{x} = A^{-1}\vec{b}$. Now we just need to compute what A^{-1} is. We can surely do elimination every time we want to find $A^{-1}\vec{b}$, but that would be ridiculous.

The mapping A^{-1} is linear and hence given by a matrix, and we have seen that to figure out the matrix we just need to find where A^{-1} takes the standard basis vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$.

That is, to find the first column of A^{-1}, we solve $A\vec{x} = \vec{e}_1$, because then $A^{-1}\vec{e}_1 = \vec{x}$. To find the second column of A^{-1}, we solve $A\vec{x} = \vec{e}_2$. And so on. It is really just n eliminations that we need to do. But it gets even easier. If you think about it, the elimination is the same for everything on the left side of the augmented matrix. Doing n eliminations separately we would redo most of the computations. Best is to do all at once.

Therefore, to find the inverse of A, we write an $n \times 2n$ augmented matrix $[\,A \mid I\,]$, where I is the identity matrix, whose columns are precisely the standard basis vectors. We then perform row reduction until we arrive at the reduced row echelon form. If A is invertible, then pivots can be found in every column of A, and so the reduced row echelon form of $[\,A \mid I\,]$ looks like $[\,I \mid A^{-1}\,]$. We then just read off the inverse A^{-1}. If you do not find a pivot in every one of the first n columns of the augmented matrix, then A is not invertible.

This is best seen by example. Suppose we wish to invert the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix}.$$

We write the augmented matrix and we start reducing:

$$\left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 0 & -4 & -5 & -2 & 1 & 0 \\ 0 & -5 & -9 & -3 & 0 & 1 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 0 & \boxed{1} & 5/4 & 1/2 & 1/4 & 0 \\ 0 & -5 & -9 & -3 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 0 & \boxed{1} & 5/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & -11/4 & -1/2 & -5/4 & 1 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 3 & 1 & 0 & 0 \\ 0 & \boxed{1} & 5/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & \boxed{1} & 2/11 & 5/11 & -4/11 \end{array}\right] \to \left[\begin{array}{ccc|ccc} \boxed{1} & 2 & 0 & 5/11 & -5/11 & 12/11 \\ 0 & \boxed{1} & 0 & 3/11 & -9/11 & 5/11 \\ 0 & 0 & \boxed{1} & 2/11 & 5/11 & -4/11 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} \boxed{1} & 0 & 0 & -1/11 & 3/11 & 2/11 \\ 0 & \boxed{1} & 0 & 3/11 & -9/11 & 5/11 \\ 0 & 0 & \boxed{1} & 2/11 & 5/11 & -4/11 \end{array}\right].$$

So

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} -1/11 & 3/11 & 2/11 \\ 3/11 & -9/11 & 5/11 \\ 2/11 & 5/11 & -4/11 \end{bmatrix}.$$

Not too terrible, no? Perhaps harder than inverting a 2×2 matrix for which we had a simple formula, but not too bad. Really in practice this is done efficiently by a computer.

A.3.6 Exercises

***Exercise* A.3.1:** *Compute the reduced row echelon form for the following matrices:*

a) $\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ *b)* $\begin{bmatrix} 3 & 3 \\ 6 & -3 \end{bmatrix}$ *c)* $\begin{bmatrix} 3 & 6 \\ -2 & -3 \end{bmatrix}$ *d)* $\begin{bmatrix} 6 & 6 & 7 & 7 \\ 1 & 1 & 0 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 9 & 3 & 0 & 2 \\ 8 & 6 & 3 & 6 \\ 7 & 9 & 7 & 9 \end{bmatrix}$ *f)* $\begin{bmatrix} 2 & 1 & 3 & -3 \\ 6 & 0 & 0 & -1 \\ -2 & 4 & 4 & 3 \end{bmatrix}$ *g)* $\begin{bmatrix} 6 & 6 & 5 \\ 0 & -2 & 2 \\ 6 & 5 & 6 \end{bmatrix}$ *h)* $\begin{bmatrix} 0 & 2 & 0 & -1 \\ 6 & 6 & -3 & 3 \\ 6 & 2 & -3 & 5 \end{bmatrix}$

***Exercise* A.3.2:** *Compute the inverse of the given matrices*

a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}$

***Exercise* A.3.3:** *Solve (find all solutions), or show no solution exists*

a) $\begin{aligned} 4x_1 + 3x_2 &= -2 \\ -x_1 + x_2 &= 4 \end{aligned}$ *b)* $\begin{aligned} x_1 + 5x_2 + 3x_3 &= 7 \\ 8x_1 + 7x_2 + 8x_3 &= 8 \\ 4x_1 + 8x_2 + 6x_3 &= 4 \end{aligned}$

c) $\begin{aligned} 4x_1 + 8x_2 + 2x_3 &= 3 \\ -x_1 - 2x_2 + 3x_3 &= 1 \\ 4x_1 + 8x_2 &= 2 \end{aligned}$ *d)* $\begin{aligned} x + 2y + 3z &= 4 \\ 2x - y + 3z &= 1 \\ 3x + y + 6z &= 6 \end{aligned}$

***Exercise* A.3.4:** *By computing the inverse, solve the following systems for $\vec{x}$.*

a) $\begin{bmatrix} 4 & 1 \\ -1 & 3 \end{bmatrix} \vec{x} = \begin{bmatrix} 13 \\ 26 \end{bmatrix}$ *b)* $\begin{bmatrix} 3 & 3 \\ 3 & 4 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

***Exercise* A.3.5:** *Compute the rank of the given matrices*

a) $\begin{bmatrix} 6 & 3 & 5 \\ 1 & 4 & 1 \\ 7 & 7 & 6 \end{bmatrix}$ *b)* $\begin{bmatrix} 5 & -2 & -1 \\ 3 & 0 & 6 \\ 2 & 4 & 5 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 2 & 4 & 6 \end{bmatrix}$

***Exercise* A.3.6:** *For the matrices in Exercise A.3.5, find a linearly independent set of row vectors that span the row space (they do not need to be rows of the matrix).*

***Exercise* A.3.7:** *For the matrices in Exercise A.3.5, find a linearly independent set of columns that span the column space. That is, find the pivot columns of the matrices.*

***Exercise* A.3.8:** *Find a linearly independent subset of the following vectors that has the same span.*

$$\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -2 \\ -4 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$$

Exercise* A.3.101*: *Compute the reduced row echelon form for the following matrices:*

a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ *c)* $\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}$ *d)* $\begin{bmatrix} 1 & -3 & 1 \\ 4 & 6 & -2 \\ -2 & 6 & -2 \end{bmatrix}$

e) $\begin{bmatrix} 2 & 2 & 5 & 2 \\ 1 & -2 & 4 & -1 \\ 0 & 3 & 1 & -2 \end{bmatrix}$ *f)* $\begin{bmatrix} -2 & 6 & 4 & 3 \\ 6 & 0 & -3 & 0 \\ 4 & 2 & -1 & 1 \end{bmatrix}$ *g)* $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *h)* $\begin{bmatrix} 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 5 \end{bmatrix}$

Exercise* A.3.102*: *Compute the inverse of the given matrices*

a) $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ *c)* $\begin{bmatrix} 2 & 4 & 0 \\ 2 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$

Exercise* A.3.103*: *Solve (find all solutions), or show no solution exists*

a) $\begin{aligned} 4x_1 + 3x_2 &= -1 \\ 5x_1 + 6x_2 &= 4 \end{aligned}$ *b)* $\begin{aligned} 5x + 6y + 5z &= 7 \\ 6x + 8y + 6z &= -1 \\ 5x + 2y + 5z &= 2 \end{aligned}$

c) $\begin{aligned} a + b + c &= -1 \\ a + 5b + 6c &= -1 \\ -2a + 5b + 6c &= 8 \end{aligned}$ *d)* $\begin{aligned} -2x_1 + 2x_2 + 8x_3 &= 6 \\ x_2 + x_3 &= 2 \\ x_1 + 4x_2 + x_3 &= 7 \end{aligned}$

Exercise* A.3.104*: *By computing the inverse, solve the following systems for $\vec{x}$.*

a) $\begin{bmatrix} -1 & 1 \\ 3 & 3 \end{bmatrix} \vec{x} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ *b)* $\begin{bmatrix} 2 & 7 \\ 1 & 6 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

Exercise* A.3.105*: *Compute the rank of the given matrices*

a) $\begin{bmatrix} 7 & -1 & 6 \\ 7 & 7 & 7 \\ 7 & 6 & 2 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$ *c)* $\begin{bmatrix} 0 & 3 & -1 \\ 6 & 3 & 1 \\ 4 & 7 & -1 \end{bmatrix}$

Exercise* A.3.106*: *For the matrices in Exercise A.3.105, find a linearly independent set of row vectors that span the row space (they do not need to be rows of the matrix).*

Exercise* A.3.107*: *For the matrices in Exercise A.3.105, find a linearly independent set of columns that span the column space. That is, find the pivot columns of the matrices.*

Exercise* A.3.108*: *Find a linearly independent subset of the following vectors that has the same span.*

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix}$$

A.4 Subspaces, dimension, and the kernel

Note: 1 lecture

A.4.1 Subspaces, basis, and dimension

We often find ourselves looking at the set of solutions of a linear equation $L\vec{x} = \vec{0}$ for some matrix L, that is, we are interested in the kernel of L. The set of all such solutions has a nice structure: It looks and acts a lot like some euclidean space $\mathbb{R}^k$.

We say that a set S of vectors in $\mathbb{R}^n$ is a *subspace* if whenever $\vec{x}$ and $\vec{y}$ are members of S and α is a scalar, then

$$\vec{x} + \vec{y}, \qquad \text{and} \qquad \alpha\vec{x}$$

are also members of S. That is, we can add and multiply by scalars and we still land in S. So every linear combination of vectors of S is still in S. That is really what a subspace is. It is a subset where we can take linear combinations and still end up being in the subset. Consequently the span of a number of vectors is automatically a subspace.

Example A.4.1: If we let $S = \mathbb{R}^n$, then this S is a subspace of $\mathbb{R}^n$. Adding any two vectors in $\mathbb{R}^n$ gets a vector in $\mathbb{R}^n$, and so does multiplying by scalars.

The set $S' = \{\vec{0}\}$, that is, the set of the zero vector by itself, is also a subspace of $\mathbb{R}^n$. There is only one vector in this subspace, so we only need to verify the definition for that one vector, and everything checks out: $\vec{0} + \vec{0} = \vec{0}$ and $\alpha\vec{0} = \vec{0}$.

The set S'' of all the vectors of the form (a, a) for any real number a, such as $(1, 1)$, $(3, 3)$, or $(-0.5, -0.5)$ is a subspace of $\mathbb{R}^2$. Adding two such vectors, say $(1, 1) + (3, 3) = (4, 4)$ again gets a vector of the same form, and so does multiplying by a scalar, say $8(1, 1) = (8, 8)$.

If S is a subspace and we can find k linearly independent vectors in S

$$\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k,$$

such that every other vector in S is a linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$, then the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is called a *basis* of S. In other words, S is the span of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$. We say that S has dimension k, and we write

$$\dim S = k.$$

Theorem A.4.1. *If $S \subset \mathbb{R}^n$ is a subspace and S is not the trivial subspace $\{\vec{0}\}$, then there exists a unique positive integer k (the dimension) and a (not unique) basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, such that every $\vec{w}$ in S can be uniquely represented by*

$$\vec{w} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_k\vec{v}_k,$$

for some scalars α_1, α_2, $\ldots$, α_k.

Just as a vector in $\mathbb{R}^k$ is represented by a k-tuple of numbers, so is a vector in a k-dimensional subspace of $\mathbb{R}^n$ represented by a k-tuple of numbers. At least once we have fixed a basis. A different basis would give a different k-tuple of numbers for the same vector.

We should reiterate that while k is unique (a subspace cannot have two different dimensions), the set of basis vectors is not at all unique. There are lots of different bases for any given subspace. Finding just the right basis for a subspace is a large part of what one does in linear algebra. In fact, that is what we spend a lot of time on in linear differential equations, although at first glance it may not seem like that is what we are doing.

Example A.4.2: The standard basis

$$\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n,$$

is a basis of $\mathbb{R}^n$, (hence the name). So as expected

$$\dim \mathbb{R}^n = n.$$

On the other hand the subspace $\{\vec{0}\}$ is of dimension 0.

The subspace S'' from Example A.4.1, that is, the set of vectors (a, a), is of dimension 1. One possible basis is simply $\{(1, 1)\}$, the single vector $(1, 1)$: every vector in S'' can be represented by $a(1, 1) = (a, a)$. Similarly another possible basis would be $\{(-1, -1)\}$. Then the vector (a, a) would be represented as $(-a)(1, 1)$.

Row and column spaces of a matrix are also examples of subspaces, as they are given as the span of vectors. We can use what we know about rank, row spaces, and column spaces from the previous section to find a basis.

Example A.4.3: In the last section, we considered the matrix

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

Using row reduction to find the pivot columns, we found

$$\text{column space of } A \left(\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 6 \\ 3 & 6 & 7 & 8 \end{bmatrix} \right) = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} \right\}.$$

What we did was we found the basis of the column space. The basis has two elements, and so the column space of A is two-dimensional. Notice that the rank of A is two.

We would have followed the same procedure if we wanted to find the basis of the subspace X spanned by

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

We would have simply formed the matrix A with these vectors as columns and repeated the computation above. The subspace X is then the column space of A.

Example A.4.4: Consider the matrix

$$L = \begin{bmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}.$$

Conveniently, the matrix is in reduced row echelon form. The matrix is of rank 3. The column space is the span of the pivot columns. It is the 3-dimensional space

$$\text{column space of } L = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \mathbb{R}^3.$$

The row space is the 3-dimensional space

$$\text{row space of } L = \operatorname{span} \left\{ \begin{bmatrix} 1 & 2 & 0 & 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 0 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 & 5 \end{bmatrix} \right\}.$$

As these vectors have 5 components, we think of the row space of L as a subspace of $\mathbb{R}^5$.

The way the dimensions worked out in the examples is not an accident. Since the number of vectors that we needed to take was always the same as the number of pivots, and the number of pivots is the rank, we get the following result.

Theorem A.4.2 (Rank)**.** *The dimension of the column space and the dimension of the row space of a matrix A are both equal to the rank of A.*

A.4.2 Kernel

The set of solutions of a linear equation $L\vec{x} = \vec{0}$, the kernel of L, is a subspace: If $\vec{x}$ and $\vec{y}$ are solutions, then

$$L(\vec{x} + \vec{y}) = L\vec{x} + L\vec{y} = \vec{0} + \vec{0} = \vec{0}, \qquad \text{and} \qquad L(\alpha\vec{x}) = \alpha L\vec{x} = \alpha\vec{0} = \vec{0}.$$

So $\vec{x} + \vec{y}$ and $\alpha\vec{x}$ are solutions. The dimension of the kernel is called the *nullity* of the matrix.

The same sort of idea governs the solutions of linear differential equations. We try to describe the kernel of a linear differential operator, and as it is a subspace, we look for a basis of this kernel. Much of this book is dedicated to finding such bases.

The kernel of a matrix is the same as the kernel of its reduced row echelon form. For a matrix in reduced row echelon form, the kernel is rather easy to find. If a vector $\vec{x}$ is applied to a matrix L, then each entry in $\vec{x}$ corresponds to a column of L, the column that the entry multiplies. To find the kernel, pick a non-pivot column make a vector that has a -1 in the entry corresponding to this non-pivot column and zeros at all the other entries corresponding to the other non-pivot columns. Then for all the entries corresponding to pivot columns make it precisely the value in the corresponding row of the non-pivot column to make the vector be a solution to $L\vec{x} = \vec{0}$. This procedure is best understood by example.

Example A.4.5: Consider

$$L = \begin{bmatrix} \boxed{1} & 2 & 0 & 0 & 3 \\ 0 & 0 & \boxed{1} & 0 & 4 \\ 0 & 0 & 0 & \boxed{1} & 5 \end{bmatrix}.$$

This matrix is in reduced row echelon form, the pivots are marked. There are two non-pivot columns, so the kernel has dimension 2, that is, it is the span of 2 vectors. Let us find the first vector. We look at the first non-pivot column, the 2^{nd} column, and we put a -1 in the 2^{nd} entry of our vector. We put a 0 in the 5^{th} entry as the 5^{th} column is also a non-pivot column:

$$\begin{bmatrix} ? \\ -1 \\ ? \\ ? \\ 0 \end{bmatrix}.$$

Let us fill the rest. When this vector hits the first row, we get a -2 and 1 times whatever the first question mark is. So make the first question mark 2. For the second and third rows, it is sufficient to make it the question marks zero. We are really filling in the non-pivot column into the remaining entries. Let us check while marking which numbers went where:

$$\begin{bmatrix} 1 & \boxed{2} & 0 & 0 & 3 \\ 0 & \boxed{0} & 1 & 0 & 4 \\ 0 & \boxed{0} & 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} \boxed{2} \\ -1 \\ \boxed{0} \\ \boxed{0} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Yay! How about the second vector. We start with

$$\begin{bmatrix} ? \\ 0 \\ ? \\ ? \\ -1. \end{bmatrix}$$

We set the first question mark to 3, the second to 4, and the third to 5. Let us check, marking things as previously,

$$\begin{bmatrix} 1 & 2 & 0 & 0 & \boxed{3} \\ 0 & 0 & 1 & 0 & \boxed{4} \\ 0 & 0 & 0 & 1 & \boxed{5} \end{bmatrix} \begin{bmatrix} \boxed{3} \\ 0 \\ \boxed{4} \\ \boxed{5} \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

There are two non-pivot columns, so we only need two vectors. We have found the basis of the kernel. So,

$$\text{kernel of } L = \operatorname{span}\left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 4 \\ 5 \\ -1 \end{bmatrix} \right\}$$

What we did in finding a basis of the kernel is we expressed all solutions of $L\vec{x} = \vec{0}$ as a linear combination of some given vectors.

The procedure to find the basis of the kernel of a matrix L:

(i) Find the reduced row echelon form of L.

(ii) Write down the basis of the kernel as above, one vector for each non-pivot column.

The rank of a matrix is the dimension of the column space, and that is the span on the pivot columns, while the kernel is the span of vectors one for each non-pivot column. So the two numbers must add to the number of columns.

Theorem A.4.3 (Rank–Nullity)**.** *If a matrix A has n columns, rank r, and nullity k (dimension of the kernel), then*

$$n = r + k.$$

The theorem is immensely useful in applications. It allows one to compute the rank r if one knows the nullity k and vice versa, without doing any extra work.

Let us consider an example application, a simple version of the so-called *Fredholm alternative*. A similar result is true for differential equations. Consider

$$A\vec{x} = \vec{b},$$

where A is a square $n \times n$ matrix. There are then two mutually exclusive possibilities:

(i) A nonzero solution $\vec{x}$ to $A\vec{x} = \vec{0}$ exists.

(ii) The equation $A\vec{x} = \vec{b}$ has a unique solution $\vec{x}$ for every $\vec{b}$.

How does the Rank–Nullity theorem come into the picture? Well, if A has a nonzero solution $\vec{x}$ to $A\vec{x} = \vec{0}$, then the nullity k is positive. But then the rank $r = n - k$ must be less than n. It means that the column space of A is of dimension less than n, so it is a subspace that does not include everything in $\mathbb{R}^n$. So $\mathbb{R}^n$ has to contain some vector $\vec{b}$ not in the column space of A. In fact, most vectors in $\mathbb{R}^n$ are not in the column space of A.

A.4.3 Exercises

Exercise* A.4.1*: *For the following sets of vectors, find a basis for the subspace spanned by the vectors, and find the dimension of the subspace.*

a) $\begin{bmatrix}1\\1\\1\end{bmatrix}$, $\begin{bmatrix}-1\\-1\\-1\end{bmatrix}$ *b)* $\begin{bmatrix}1\\0\\5\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\end{bmatrix}$, $\begin{bmatrix}0\\-1\\0\end{bmatrix}$ *c)* $\begin{bmatrix}-4\\-3\\5\end{bmatrix}$, $\begin{bmatrix}2\\3\\3\end{bmatrix}$, $\begin{bmatrix}2\\0\\2\end{bmatrix}$

d) $\begin{bmatrix}1\\3\\0\end{bmatrix}$, $\begin{bmatrix}0\\2\\2\end{bmatrix}$, $\begin{bmatrix}-1\\-1\\2\end{bmatrix}$ *e)* $\begin{bmatrix}1\\3\end{bmatrix}$, $\begin{bmatrix}0\\2\end{bmatrix}$, $\begin{bmatrix}-1\\-1\end{bmatrix}$ *f)* $\begin{bmatrix}3\\1\\3\end{bmatrix}$, $\begin{bmatrix}2\\4\\-4\end{bmatrix}$, $\begin{bmatrix}-5\\-5\\-2\end{bmatrix}$

Exercise* A.4.2*: *For the following matrices, find a basis for the kernel (nullspace).*

a) $\begin{bmatrix}1&1&1\\1&1&5\\1&1&-4\end{bmatrix}$ *b)* $\begin{bmatrix}2&-1&-3\\4&0&-4\\-1&1&2\end{bmatrix}$ *c)* $\begin{bmatrix}-4&4&4\\-1&1&1\\-5&5&5\end{bmatrix}$ *d)* $\begin{bmatrix}-2&1&1&1\\-4&2&2&2\\1&0&4&3\end{bmatrix}$

Exercise* A.4.3*: *Suppose a* 5×5 *matrix* A *has rank 3. What is the nullity?*

Exercise* A.4.4*: *Suppose that* X *is the set of all the vectors of* $\mathbb{R}^3$ *whose third component is zero. Is* X *a subspace? And if so, find a basis and the dimension.*

Exercise* A.4.5*: *Consider a square matrix* A*, and suppose that* $\vec{x}$ *is a nonzero vector such that* $A\vec{x} = \vec{0}$*. What does the Fredholm alternative say about invertibility of* A*.*

Exercise* A.4.6*: *Consider*

$$M = \begin{bmatrix}1&2&3\\2&?&?\\-1&?&?\end{bmatrix}.$$

If the nullity of this matrix is 2, fill in the question marks. Hint: What is the rank?

Exercise* A.4.101*: *For the following sets of vectors, find a basis for the subspace spanned by the vectors, and find the dimension of the subspace.*

a) $\begin{bmatrix}1\\2\end{bmatrix}$, $\begin{bmatrix}1\\1\end{bmatrix}$ *b)* $\begin{bmatrix}1\\1\\1\end{bmatrix}$, $\begin{bmatrix}2\\2\\2\end{bmatrix}$, $\begin{bmatrix}1\\1\\2\end{bmatrix}$ *c)* $\begin{bmatrix}5\\3\\1\end{bmatrix}$, $\begin{bmatrix}5\\-1\\5\end{bmatrix}$, $\begin{bmatrix}-1\\3\\-4\end{bmatrix}$

d) $\begin{bmatrix}2\\2\\4\end{bmatrix}$, $\begin{bmatrix}2\\2\\3\end{bmatrix}$, $\begin{bmatrix}4\\4\\-3\end{bmatrix}$ *e)* $\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}2\\0\end{bmatrix}$, $\begin{bmatrix}3\\0\end{bmatrix}$ *f)* $\begin{bmatrix}1\\0\\0\end{bmatrix}$, $\begin{bmatrix}2\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\2\end{bmatrix}$

Exercise* A.4.102*: *For the following matrices, find a basis for the kernel (nullspace).*

a) $\begin{bmatrix}2&6&1&9\\1&3&2&9\\3&9&0&9\end{bmatrix}$ *b)* $\begin{bmatrix}2&-2&-5\\-1&1&5\\-5&5&-3\end{bmatrix}$ *c)* $\begin{bmatrix}1&-5&-4\\2&3&5\\-3&5&2\end{bmatrix}$ *d)* $\begin{bmatrix}0&4&4\\0&1&1\\0&5&5\end{bmatrix}$

***Exercise* A.4.103:** *Suppose the column space of a* 9×5 *matrix* A *of dimension 3. Find*

a) *Rank of* A.

b) *Nullity of* A.

c) *Dimension of the row space of* A.

d) *Dimension of the nullspace of* A.

e) *Size of the maximum subset of linearly independent rows of* A.

A.5 Inner product and projections

Note: 1–2 lectures

A.5.1 Inner product and orthogonality

To do basic geometry, we need length, and we need angles. We have already seen the euclidean length, so let us figure out how to compute angles. Mostly, we are worried about the right angle*.

Given two (column) vectors in $\mathbb{R}^n$, we define the (standard) *inner product* as the dot product:

$$\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \vec{y}^T \vec{x} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i .$$

Why do we seemingly give a new notation for the dot product? Because there are other possible inner products, which are not the dot product, although we will not worry about others here. An inner product can even be defined on spaces of functions as we do in chapter **??**:

$$\langle f(t), g(t) \rangle = \int_a^b f(t) g(t) \, dt .$$

But we digress.

The inner product satisfies the following rules:

(i) $\langle \vec{x}, \vec{x} \rangle \geq 0$, and $\langle \vec{x}, \vec{x} \rangle = 0$ if and only if $\vec{x} = 0$,

(ii) $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$,

(iii) $\langle a\vec{x}, \vec{y} \rangle = \langle \vec{x}, a\vec{y} \rangle = a \langle \vec{x}, \vec{y} \rangle$,

(iv) $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle$ and $\langle \vec{x}, \vec{y} + \vec{z} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{x}, \vec{z} \rangle$.

Anything that satisfies the properties above can be called an inner product, although in this section we are concerned with the standard inner product in $\mathbb{R}^n$.

The standard inner product gives the euclidean length:

$$\lVert \vec{x} \rVert = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} .$$

How does it give angles?

You may recall from multivariable calculus, that in two or three dimensions, the standard inner product (the dot product) gives you the angle between the vectors:

$$\langle \vec{x}, \vec{y} \rangle = \lVert \vec{x} \rVert \lVert \vec{y} \rVert \cos \theta .$$

*When Euclid defined angles in his *Elements*, the only angle he ever really defined was the right angle.

That is, θ is the angle that $\vec{x}$ and $\vec{y}$ make when they are based at the same point.

In $\mathbb{R}^n$ (any dimension), we are simply going to say that θ from the formula is what the angle is. This makes sense as any two vectors based at the origin lie in a 2-dimensional plane (subspace), and the formula works in 2 dimensions. In fact, one could even talk about angles between functions this way, and we do in chapter **??**, where we talk about orthogonal functions (functions at right angle to each other).

To compute the angle we compute

$$\cos\theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\|\|\vec{y}\|}.$$

Our angles are always in radians. We are computing the cosine of the angle, which is really the best we can do. Given two vectors at an angle θ, we can give the angle as $-\theta$, $2\pi - \theta$, etc., see Figure A.5. Fortunately, $\cos\theta = \cos(-\theta) = \cos(2\pi - \theta)$. If we solve for θ using the inverse cosine $\cos^{-1}$, we can just decree that $0 \leq \theta \leq \pi$.

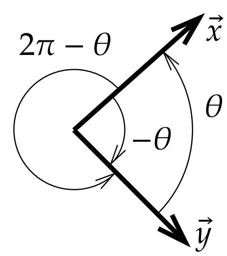


Figure A.5: Angle between vectors.

Example A.5.1: Let us compute the angle between the vectors $(3, 0)$ and $(1, 1)$ in the plane. Compute

$$\cos\theta = \frac{\big\langle (3,0), (1,1) \big\rangle}{\|(3,0)\|\|(1,1)\|} = \frac{3+0}{3\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

Therefore $\theta = \pi/4$.

As we said, the most important angle is the right angle. A right angle is $\pi/2$ radians, and $\cos(\pi/2) = 0$, so the formula is particularly easy in this case. We say vectors $\vec{x}$ and $\vec{y}$ are *orthogonal* if they are at right angles, that is if

$$\langle \vec{x}, \vec{y} \rangle = 0.$$

The vectors $(1, 0, 0, 1)$ and $(1, 2, 3, -1)$ are orthogonal. So are $(1, 1)$ and $(1, -1)$. However, $(1, 1)$ and $(1, 2)$ are not orthogonal as their inner product is 3 and not 0.

A.5.2 Orthogonal projection

A typical application of linear algebra is to take a difficult problem, write everything in the right basis, and in this new basis the problem becomes simple. A particularly useful basis is an orthogonal basis, that is a basis where all the basis vectors are orthogonal. When we draw a coordinate system in two or three dimensions, we almost always draw our axes as orthogonal to each other.

Generalizing this concept to functions, it is particularly useful in chapter **??** to express a function using a particular orthogonal basis, the Fourier series.

To express one vector in terms of an orthogonal basis, we need to first *project* one vector onto another. Given a nonzero vector $\vec{v}$, we define the *orthogonal projection* of $\vec{w}$ onto $\vec{v}$ as

$$\operatorname{proj}_{\vec{v}}(\vec{w}) = \left(\frac{\langle \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \right) \vec{v}.$$

For the geometric idea, see Figure A.6. That is, we find the "shadow of $\vec{w}$" on the line spanned by $\vec{v}$ if the direction of the sun's rays were exactly perpendicular to the line. Another way of thinking about it is that the tip of the arrow of $\operatorname{proj}_{\vec{v}}(\vec{w})$ is the closest point on the line spanned by $\vec{v}$ to the tip of the arrow of $\vec{w}$. In terms of euclidean distance, $\vec{u} = \operatorname{proj}_{\vec{v}}(\vec{w})$ minimizes the distance $\lVert \vec{w} - \vec{u} \rVert$ among all vectors $\vec{u}$ that are multiples of $\vec{v}$. Because of this, this projection comes up often in applied mathematics in all sorts of contexts we cannot solve a problem exactly: We cannot always solve "Find $\vec{w}$ as a multiple of $\vec{v}$," but $\operatorname{proj}_{\vec{v}}(\vec{w})$ is the best "solution."

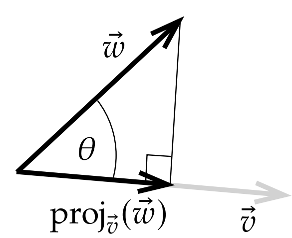


Figure A.6: Orthogonal projection.

The formula follows from basic trigonometry. The length of $\operatorname{proj}_{\vec{v}}(\vec{w})$ should be $\cos\theta$ times the length of $\vec{w}$, that is $(\cos\theta)\lVert \vec{w} \rVert$. We take the unit vector in the direction of $\vec{v}$, that is, $\frac{\vec{v}}{\lVert \vec{v} \rVert}$ and we multiply it by the length of the projection. In other words,

$$\operatorname{proj}_{\vec{v}}(\vec{w}) = (\cos\theta)\lVert \vec{w} \rVert \frac{\vec{v}}{\lVert \vec{v} \rVert} = \frac{(\cos\theta)\lVert \vec{w} \rVert \lVert \vec{v} \rVert}{\lVert \vec{v} \rVert^2} \vec{v} = \frac{\langle \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}.$$

Example A.5.2: Suppose we wish to project the vector $(3, 2, 1)$ onto the vector $(1, 2, 3)$.

Compute

$$\operatorname{proj}_{(1,2,3)}\big((3,2,1)\big) = \frac{\langle (3,2,1),(1,2,3)\rangle}{\langle (1,2,3),(1,2,3)\rangle}(1,2,3) = \frac{3\cdot 1 + 2\cdot 2 + 1\cdot 3}{1\cdot 1 + 2\cdot 2 + 3\cdot 3}(1,2,3)$$
$$= \frac{10}{14}(1,2,3) = \left(\frac{5}{7},\frac{10}{7},\frac{15}{7}\right).$$

Let us double check that the projection is orthogonal. That is $\vec{w} - \operatorname{proj}_{\vec{v}}(\vec{w})$ ought to be orthogonal to $\vec{v}$, see the right angle in Figure A.6 on the facing page. That is,

$$(3,2,1) - \operatorname{proj}_{(1,2,3)}\big((3,2,1)\big) = \left(3-\frac{5}{7}, 2-\frac{10}{7}, 1-\frac{15}{7}\right) = \left(\frac{16}{7},\frac{4}{7},\frac{-8}{7}\right)$$

ought to be orthogonal to $(1,2,3)$. We compute the inner product and we had better get zero:

$$\left\langle \left(\frac{16}{7},\frac{4}{7},\frac{-8}{7}\right), (1,2,3)\right\rangle = \frac{16}{7}\cdot 1 + \frac{4}{7}\cdot 2 - \frac{8}{7}\cdot 3 = 0.$$

A.5.3 Orthogonal basis

As we said, a basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is an *orthogonal basis* if all vectors in the basis are orthogonal to each other, that is, if

$$\langle \vec{v}_j, \vec{v}_k \rangle = 0$$

for all choices of j and k where $j \neq k$ (a nonzero vector cannot be orthogonal to itself). A basis is furthermore called an *orthonormal basis* if all the vectors in a basis are also unit vectors, that is, if all the vectors have magnitude 1. For example, the standard basis $\{(1,0,0),(0,1,0),(0,0,1)\}$ is an orthonormal basis of $\mathbb{R}^3$: Any pair is orthogonal, and each vector is of unit magnitude.

The reason why we are interested in orthogonal (or orthonormal) bases is that they make it really simple to represent a vector (or a projection onto a subspace) in the basis. The simple formula for the orthogonal projection onto a vector gives us the coefficients. In chapter **??**, we use the same idea by finding the correct orthogonal basis for the set of solutions of a differential equation. We are then able to find any particular solution by simply applying the orthogonal projection formula, which is just a couple of a inner products.

Let us come back to linear algebra. Suppose that we have a subspace and an orthogonal basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. We wish to express $\vec{x}$ in terms of the basis. If $\vec{x}$ is not in the span of the basis (when it is not in the given subspace), then of course it is not possible, but the following formula gives us at least the orthogonal projection onto the subspace, or in other words, the best approximation in the subspace.

First suppose that $\vec{x}$ is in the span. Then it is the sum of the orthogonal projections:

$$\vec{x} = \operatorname{proj}_{\vec{v}_1}(\vec{x}) + \operatorname{proj}_{\vec{v}_2}(\vec{x}) + \cdots + \operatorname{proj}_{\vec{v}_n}(\vec{x}) = \frac{\langle \vec{x}, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle}\vec{v}_1 + \frac{\langle \vec{x}, \vec{v}_2\rangle}{\langle \vec{v}_2, \vec{v}_2\rangle}\vec{v}_2 + \cdots + \frac{\langle \vec{x}, \vec{v}_n\rangle}{\langle \vec{v}_n, \vec{v}_n\rangle}\vec{v}_n.$$

In other words, if we want to write $\vec{x} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$, then

$$a_1 = \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}, \quad a_2 = \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}, \quad \ldots, \quad a_n = \frac{\langle \vec{x}, \vec{v}_n \rangle}{\langle \vec{v}_n, \vec{v}_n \rangle}.$$

Another way to derive this formula is to work in reverse. Suppose that $\vec{x} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$. Take an inner product with $\vec{v}_j$, and use the properties of the inner product:

$$\begin{aligned} \langle \vec{x}, \vec{v}_j \rangle &= \langle a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n, \vec{v}_j \rangle \\ &= a_1\langle \vec{v}_1, \vec{v}_j \rangle + a_2\langle \vec{v}_2, \vec{v}_j \rangle + \cdots + a_n\langle \vec{v}_n, \vec{v}_j \rangle. \end{aligned}$$

As the basis is orthogonal, then $\langle \vec{v}_k, \vec{v}_j \rangle = 0$ whenever $k \neq j$. That means that only one of the terms, the j^{th} one, on the right-hand side is nonzero and we get

$$\langle \vec{x}, \vec{v}_j \rangle = a_j \langle \vec{v}_j, \vec{v}_j \rangle.$$

Solving for a_j we find $a_j = \frac{\langle \vec{x}, \vec{v}_j \rangle}{\langle \vec{v}_j, \vec{v}_j \rangle}$ as before.

Example A.5.3: The vectors $(1, 1)$ and $(1, -1)$ form an orthogonal basis of $\mathbb{R}^2$. Suppose we wish to represent $(3, 4)$ in terms of this basis, that is, we wish to find a_1 and a_2 such that

$$(3, 4) = a_1(1, 1) + a_2(1, -1).$$

We compute:

$$a_1 = \frac{\langle (3,4), (1,1) \rangle}{\langle (1,1), (1,1) \rangle} = \frac{7}{2}, \qquad a_2 = \frac{\langle (3,4), (1,-1) \rangle}{\langle (1,-1), (1,-1) \rangle} = \frac{-1}{2}.$$

So

$$(3, 4) = \frac{7}{2}(1, 1) + \frac{-1}{2}(1, -1).$$

If the basis is orthonormal rather than orthogonal, then all the denominators are one. It is easy to make a basis orthonormal—divide all the vectors by their size. If you want to decompose many vectors, it may be better to find an orthonormal basis. In the example above, the orthonormal basis we would thus create is

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \quad \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right).$$

Then the computation would have been

$$\begin{aligned} (3, 4) &= \left\langle (3,4), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right\rangle \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) + \left\langle (3,4), \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) \right\rangle \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) \\ &= \frac{7}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) + \frac{-1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right). \end{aligned}$$

Maybe the example is not so awe inspiring, but given vectors in $\mathbb{R}^{20}$ rather than $\mathbb{R}^2$, then surely one would much rather do 20 inner products (or 40 if we did not have an orthonormal basis) rather than solving a system of twenty equations in twenty unknowns using row reduction of a 20×21 matrix.

As we said above, the formula still works even if $\vec{x}$ is not in the subspace, although then it does not get us the vector $\vec{x}$ but its projection. More concretely, suppose that S is a subspace that is the span of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ and $\vec{x}$ is any vector. Let $\operatorname{proj}_S(\vec{x})$ be the vector in S that is the closest to $\vec{x}$. Then

$$\operatorname{proj}_S(\vec{x}) = \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 + \cdots + \frac{\langle \vec{x}, \vec{v}_n \rangle}{\langle \vec{v}_n, \vec{v}_n \rangle} \vec{v}_n .$$

Of course, if $\vec{x}$ is in S, then $\operatorname{proj}_S(\vec{x}) = \vec{x}$, as the closest vector in S to $\vec{x}$ is $\vec{x}$ itself. But true utility is obtained when $\vec{x}$ is not in S. In much of applied mathematics, we cannot find an exact solution to a problem, but we try to find the best solution out of a small subset (subspace). The partial sums of Fourier series from chapter **??** are one example. Another example is least square approximation to fit a curve to data. Yet another example is given by the most commonly used numerical methods to solve partial differential equations, the finite element methods.

Example A.5.4: The vectors $(1,2,3)$ and $(3,0,-1)$ are orthogonal, and so they are an orthogonal basis of a subspace S:

$$S = \operatorname{span}\bigl\{ (1,2,3), (3,0,-1) \bigr\}.$$

Let us find the vector in S that is closest to $(2,1,0)$. That is, let us find $\operatorname{proj}_S\bigl((2,1,0)\bigr)$.

$$\begin{aligned}
\operatorname{proj}_S\bigl((2,1,0)\bigr) &= \frac{\langle (2,1,0), (1,2,3) \rangle}{\langle (1,2,3), (1,2,3) \rangle} (1,2,3) + \frac{\langle (2,1,0), (3,0,-1) \rangle}{\langle (3,0,-1), (3,0,-1) \rangle} (3,0,-1) \\
&= \frac{2}{7}(1,2,3) + \frac{3}{5}(3,0,-1) \\
&= \left(\frac{73}{35}, \frac{4}{7}, \frac{9}{35} \right).
\end{aligned}$$

A.5.4 The Gram–Schmidt process

Before leaving orthogonal bases, let us note a procedure for manufacturing them out of any old basis. It may not be difficult to come up with an orthogonal basis for a 2-dimensional subspace, but for a 20-dimensional subspace, it seems a daunting task. Fortunately, the orthogonal projection can be used to “project away” the bits of the vectors that are making them not orthogonal. It is called the *Gram–Schmidt process*.

We start with a basis of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. We construct an orthogonal basis $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n$ as follows.

$$\vec{w}_1 = \vec{v}_1,$$

$$
\begin{aligned}
\vec{w}_2 &= \vec{v}_2 - \operatorname{proj}_{\vec{w}_1}(\vec{v}_2), \\
\vec{w}_3 &= \vec{v}_3 - \operatorname{proj}_{\vec{w}_1}(\vec{v}_3) - \operatorname{proj}_{\vec{w}_2}(\vec{v}_3), \\
\vec{w}_4 &= \vec{v}_4 - \operatorname{proj}_{\vec{w}_1}(\vec{v}_4) - \operatorname{proj}_{\vec{w}_2}(\vec{v}_4) - \operatorname{proj}_{\vec{w}_3}(\vec{v}_4), \\
&\vdots \\
\vec{w}_n &= \vec{v}_n - \operatorname{proj}_{\vec{w}_1}(\vec{v}_n) - \operatorname{proj}_{\vec{w}_2}(\vec{v}_n) - \cdots - \operatorname{proj}_{\vec{w}_{n-1}}(\vec{v}_n).
\end{aligned}
$$

What we do is at the k^{th} step, we take $\vec{v}_k$ and we subtract the projection of $\vec{v}_k$ to the subspace spanned by $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_{k-1}$.

Example A.5.5: Consider the vectors $(1,2,-1)$, and $(0,5,-2)$ and call S the span of the two vectors. Let us find an orthogonal basis of S:

$$
\begin{aligned}
\vec{w}_1 &= (1,2,-1), \\
\vec{w}_2 &= (0,5,-2) - \operatorname{proj}_{(1,2,-1)}\big((0,2,-2)\big) \\
&= (0,1,-1) - \frac{\langle (0,5,-2), (1,2,-1) \rangle}{\langle (1,2,-1), (1,2,-1) \rangle}(1,2,-1) = (0,5,-2) - 2(1,2,-1) = (-2,1,0).
\end{aligned}
$$

So $(1,2,-1)$ and $(-2,1,0)$ span S and are orthogonal. Let us check: $\langle (1,2,-1), (-2,1,0) \rangle = 0$.

Suppose we wish to find an orthonormal basis, not just an orthogonal one. Well, we simply make the vectors into unit vectors by dividing them by their magnitude. The two vectors making up the orthonormal basis of S are:

$$
\frac{1}{\sqrt{6}}(1,2,-1) = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{-1}{\sqrt{6}} \right), \qquad \frac{1}{\sqrt{5}}(-2,1,0) = \left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0 \right).
$$

A.5.5 Exercises

***Exercise* A.5.1:** *Find the s that makes the following vectors orthogonal:* $(1,2,3)$, $(1,1,s)$.

***Exercise* A.5.2:** *Find the angle θ between* $(1,3,1)$, $(2,1,-1)$.

***Exercise* A.5.3:** *Given that* $\langle \vec{v}, \vec{w} \rangle = 3$ *and* $\langle \vec{v}, \vec{u} \rangle = -1$ *compute*

a) $\langle \vec{u}, 2\vec{v} \rangle$ *b)* $\langle \vec{v}, 2\vec{w} + 3\vec{u} \rangle$ *c)* $\langle \vec{w} + 3\vec{u}, \vec{v} \rangle$

***Exercise* A.5.4:** *Suppose* $\vec{v} = (1,1,-1)$. *Find*

a) $\operatorname{proj}_{\vec{v}}\big((1,0,0)\big)$ *b)* $\operatorname{proj}_{\vec{v}}\big((1,2,3)\big)$ *c)* $\operatorname{proj}_{\vec{v}}\big((1,-1,0)\big)$

***Exercise* A.5.5:** *Consider the vectors* $(1,2,3)$, $(-3,0,1)$, $(1,-5,3)$.

a) Check that the vectors are linearly independent and so form a basis.

b) Check that the vectors are mutually orthogonal, and are therefore an orthogonal basis.

c) Represent $(1,1,1)$ *as a linear combination of this basis.*

d) Make the basis orthonormal.

Exercise* A.5.6:* *Let S be the subspace spanned by* $(1, 3, -1)$, $(1, 1, 1)$. *Find an orthogonal basis of S by the Gram-Schmidt process.*

Exercise* A.5.7:* *Starting with* $(1, 2, 3)$, $(1, 1, 1)$, $(2, 2, 0)$, *follow the Gram-Schmidt process to find an orthogonal basis of* $\mathbb{R}^3$.

Exercise* A.5.8:* *Find an orthogonal basis of* $\mathbb{R}^3$ *such that* $(3, 1, -2)$ *is one of the vectors. Hint: First find two extra vectors to make a linearly independent set.*

Exercise* A.5.9:* *Using cosines and sines of θ, find a unit vector $\vec{u}$ in $\mathbb{R}^2$ that makes angle θ with* $\vec{\imath} = (1, 0)$. *What is* $\langle \vec{\imath}, \vec{u} \rangle$*?*

Exercise* A.5.101:* *Find the s that makes the following vectors orthogonal:* $(1, 1, 1)$, $(1, s, 1)$.

Exercise* A.5.102:* *Find the angle θ between* $(1, 2, 3)$, $(1, 1, 1)$.

Exercise* A.5.103:* *Given that* $\langle \vec{v}, \vec{w} \rangle = 1$ *and* $\langle \vec{v}, \vec{u} \rangle = -1$ *and* $\|\vec{v}\| = 3$ *and*

a) $\langle 3\vec{u}, 5\vec{v} \rangle$ *b)* $\langle \vec{v}, 2\vec{w} + 3\vec{u} \rangle$ *c)* $\langle \vec{w} + 3\vec{v}, \vec{v} \rangle$

Exercise* A.5.104:* *Suppose* $\vec{v} = (1, 0, -1)$. *Find*

a) $\operatorname{proj}_{\vec{v}}\big((0, 2, 1)\big)$ *b)* $\operatorname{proj}_{\vec{v}}\big((1, 0, 1)\big)$ *c)* $\operatorname{proj}_{\vec{v}}\big((4, -1, 0)\big)$

Exercise* A.5.105:* *The vectors* $(1, 1, -1)$, $(2, -1, 1)$, $(1, -5, 3)$ *form an orthogonal basis. Represent the following vectors in terms of this basis:*

a) $(1, -8, 4)$ *b)* $(5, -7, 5)$ *c)* $(0, -6, 2)$

Exercise* A.5.106:* *Let S be the subspace spanned by* $(2, -1, 1)$, $(2, 2, 2)$. *Find an orthogonal basis of S by the Gram-Schmidt process.*

Exercise* A.5.107:* *Starting with* $(1, 1, -1)$, $(2, 3, -1)$, $(1, -1, 1)$, *follow the Gram-Schmidt process to find an orthogonal basis of* $\mathbb{R}^3$.

A.6 Determinant

Note: 1 lecture

For square matrices we define a useful quantity called the *determinant*. Define the determinant of a 1×1 matrix as the value of its only entry

$$\det\left(\begin{bmatrix} a \end{bmatrix}\right) \overset{\text{def}}{=} a.$$

For a 2×2 matrix, define

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \overset{\text{def}}{=} ad - bc.$$

Before defining the determinant for larger matrices, we note the meaning of the determinant. An $n \times n$ matrix gives a mapping of the n-dimensional euclidean space $\mathbb{R}^n$ to itself. So a 2×2 matrix A is a mapping of the plane to itself. The determinant of A is the factor by which the area of objects changes. If we take the unit square (square of side 1) in the plane, then A takes the square to a parallelogram of area $|\det(A)|$. The sign of $\det(A)$ denotes a change of orientation (negative if the axes get flipped). For example, let

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

Then $\det(A) = 1 + 1 = 2$. Let us see where A sends the unit square—the square with vertices $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$. The point $(0,0)$ gets sent to $(0,0)$.

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

The image of the square is another square with vertices $(0,0)$, $(1,-1)$, $(1,1)$, and $(2,0)$. The image square has a side of length $\sqrt{2}$, and it is therefore of area 2. See Figure A.7.

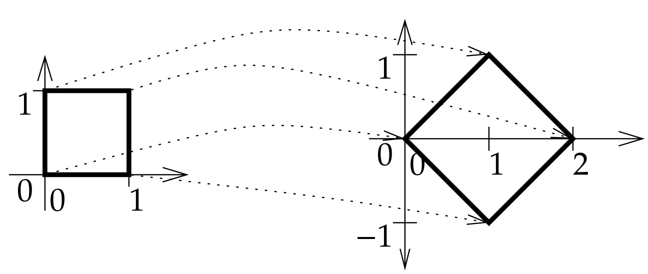


Figure A.7: Image of the unit quare via the mapping A.

In general, the image of a square is going to be a parallelogram. In high school geometry, you may have seen a formula for computing the area of a parallelogram with vertices $(0,0)$,

(a, c), (b, d) and $(a + b, c + d)$. The area is

$$\left|\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)\right| = |ad - bc|.$$

The vertical lines above mean absolute value. The matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ carries the unit square to the given parallelogram.

There are a number of ways to define the determinant for an $n \times n$ matrix. Let us use the so-called *cofactor expansion*. We define A_{ij} as the matrix A with the i^{th} row and the j^{th} column deleted. For example, if

$$\text{If} \qquad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \qquad \text{then} \qquad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} \qquad \text{and} \qquad A_{23} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}.$$

We now define the determinant recursively

$$\det(A) \overset{\text{def}}{=} \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}),$$

or in other words

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) - \cdots \begin{cases} +a_{1n}\det(A_{1n}) & \text{if } n \text{ is odd,} \\ -a_{1n}\det(A_{1n}) & \text{if } n \text{ even.} \end{cases}$$

For a 3×3 matrix, we get $\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13})$. For example,

$$\begin{aligned} \det\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right) &= 1 \cdot \det\left(\begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}\right) - 2 \cdot \det\left(\begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}\right) + 3 \cdot \det\left(\begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}\right) \\ &= 1(5 \cdot 9 - 6 \cdot 8) - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) = 0. \end{aligned}$$

It turns out that we did not have to necessarily use the first row. That is for any i,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

It is sometimes useful to use a row other than the first. In the following example it is more convenient to expand along the second row. Notice that for the second row we are starting with a negative sign.

$$\begin{aligned} \det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 0 \\ 7 & 8 & 9 \end{bmatrix}\right) &= -0 \cdot \det\left(\begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix}\right) + 5 \cdot \det\left(\begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix}\right) - 0 \cdot \det\left(\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}\right) \\ &= 0 + 5(1 \cdot 9 - 3 \cdot 7) + 0 = -60. \end{aligned}$$

Let us check if it is really the same as expanding along the first row,

$$\det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 0 \\ 7 & 8 & 9 \end{bmatrix}\right) = 1 \cdot \det\left(\begin{bmatrix} 5 & 0 \\ 8 & 9 \end{bmatrix}\right) - 2 \cdot \det\left(\begin{bmatrix} 0 & 0 \\ 7 & 9 \end{bmatrix}\right) + 3 \cdot \det\left(\begin{bmatrix} 0 & 5 \\ 7 & 8 \end{bmatrix}\right)$$
$$= 1(5 \cdot 9 - 0 \cdot 8) - 2(0 \cdot 9 - 0 \cdot 7) + 3(0 \cdot 8 - 5 \cdot 7) = -60.$$

In computing the determinant, we alternately add and subtract the determinants of the submatrices A_{ij} multiplied by a_{ij} for a fixed i and all j. The numbers $(-1)^{i+j}\det(A_{ij})$ are called *cofactors* of the matrix. And that is why this method of computing the determinant is called the *cofactor expansion*.

Similarly we do not need to expand along a row, we can expand along a column. For any j,

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

A related fact is that

$$\det(A) = \det(A^T).$$

A matrix is *upper triangular* if all elements below the main diagonal are 0. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$$

is upper triangular. Similarly a *lower triangular* matrix is one where everything above the diagonal is zero. For example,

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

The determinant for triangular matrices is very simple to compute. Consider the lower triangular matrix. If we expand along the first row, we find that the determinant is 1 times the determinant of the lower triangular matrix $\left[\begin{smallmatrix} 5 & 0 \\ 8 & 9 \end{smallmatrix}\right]$. So the deteriminant is just the product of the diagonal entries:

$$\det\left(\begin{bmatrix} 1 & 0 & 0 \\ 4 & 5 & 0 \\ 7 & 8 & 9 \end{bmatrix}\right) = 1 \cdot 5 \cdot 9 = 45.$$

Similarly for upper triangular matrices

$$\det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}\right) = 1 \cdot 5 \cdot 9 = 45.$$

In general, if A is triangular, then

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

If A is diagonal, then it is also triangular (upper and lower), so same formula applies. For example,

$$\det\left(\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}\right) = 2 \cdot 3 \cdot 5 = 30.$$

In particular, the identity matrix I is diagonal, and the diagonal entries are all 1. Thus,

$$\det(I) = 1.$$

The determinant is telling you how geometric objects scale. If B doubles the sizes of geometric objects and A triples them, then AB (which applies B to an object and then it applies A) should make size go up by a factor of 6. This is true in general:

Theorem A.6.1.

$$\det(AB) = \det(A)\det(B).$$

This property is one of the most useful, and it is employed often to actually compute determinants. A particularly interesting consequence is to note what it means for the existence of inverses. Take A and B to be inverses, that is $AB = I$. Then

$$\det(A)\det(B) = \det(AB) = \det(I) = 1.$$

Neither $\det(A)$ nor $\det(B)$ can be zero. This fact is an extremely useful property of the determinant, and one which is used often in this book:

Theorem A.6.2. *An $n \times n$ matrix A is invertible if and only if $\det(A) \neq 0$.*

In fact, $\det(A^{-1})\det(A) = 1$ says that

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

So we know what the determinant of A^{-1} is without computing A^{-1}.

Let us return to the formula for the inverse of a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Notice the determinant of the matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ in the denominator of the fraction. The formula only works if the determinant is nonzero, otherwise we are dividing by zero.

A common notation for the determinant is a pair of vertical lines:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

Personally, I find this notation confusing as vertical lines usually mean a positive quantity, while determinants can be negative. Also think about how to write the absolute value of a determinant. This notation is not used in this book.

A.6.1 Exercises

Exercise* A.6.1*: *Compute the determinant of the following matrices:*

a) $\begin{bmatrix} 3 \end{bmatrix}$ *b)* $\begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$ *c)* $\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ *d)* $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$

e) $\begin{bmatrix} 2 & 1 & 0 \\ -2 & 7 & -3 \\ 0 & 2 & 0 \end{bmatrix}$ *f)* $\begin{bmatrix} 2 & 1 & 3 \\ 8 & 6 & 3 \\ 7 & 9 & 7 \end{bmatrix}$ *g)* $\begin{bmatrix} 0 & 2 & 5 & 7 \\ 0 & 0 & 2 & -3 \\ 3 & 4 & 5 & 7 \\ 0 & 0 & 2 & 4 \end{bmatrix}$ *h)* $\begin{bmatrix} 0 & 1 & 2 & 0 \\ 1 & 1 & -1 & 2 \\ 1 & 1 & 2 & 1 \\ 2 & -1 & -2 & 3 \end{bmatrix}$

Exercise* A.6.2*: *For which x are the following matrices singular (not invertible).*

a) $\begin{bmatrix} 2 & 3 \\ 2 & x \end{bmatrix}$ *b)* $\begin{bmatrix} 2 & x \\ 1 & 2 \end{bmatrix}$ *c)* $\begin{bmatrix} x & 1 \\ 4 & x \end{bmatrix}$ *d)* $\begin{bmatrix} x & 0 & 1 \\ 1 & 4 & 2 \\ 1 & 6 & 2 \end{bmatrix}$

Exercise* A.6.3*: *Compute*

$$\det\left(\begin{bmatrix} 2 & 1 & 2 & 3 \\ 0 & 8 & 6 & 5 \\ 0 & 0 & 3 & 9 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}\right)$$

without computing the inverse.

Exercise* A.6.4*: *Suppose*

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 7 & \pi & 1 & 0 \\ 2^8 & 5 & -99 & 1 \end{bmatrix} \qquad \text{and} \qquad U = \begin{bmatrix} 5 & 9 & 1 & -\sin(1) \\ 0 & 1 & 88 & -1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let $A = LU$. Compute $\det(A)$ *in a simple way, without computing what is A. Hint: First read off* $\det(L)$ *and* $\det(U)$*.*

Exercise* A.6.5*: *Consider the linear mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by the matrix $A = \left[\begin{smallmatrix} 1 & x \\ 2 & 1 \end{smallmatrix}\right]$ for some number x. You wish to make A such that it doubles the area of every geometric figure. What are the possibilities for x (there are two answers).*

Exercise* A.6.6*: *Suppose A and S are $n \times n$ matrices, and S is invertible. Suppose that* $\det(A) = 3$*. Compute* $\det(S^{-1}AS)$ *and* $\det(SAS^{-1})$*. Justify your answer using the theorems in this section.*

Exercise* A.6.7*: *Let A be an $n \times n$ matrix such that* $\det(A) = 1$*. Compute* $\det(xA)$ *given a number x. Hint: First try computing* $\det(xI)$*, then note that $xA = (xI)A$.*

Exercise* A.6.101*: *Compute the determinant of the following matrices:*

a) $\begin{bmatrix} -2 \end{bmatrix}$ *b)* $\begin{bmatrix} 2 & -2 \\ 1 & 3 \end{bmatrix}$ *c)* $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ *d)* $\begin{bmatrix} 2 & 9 & -11 \\ 0 & -1 & 5 \\ 0 & 0 & 3 \end{bmatrix}$

e) $\begin{bmatrix} 2 & 1 & 0 \\ -2 & 7 & 3 \\ 1 & 1 & 0 \end{bmatrix}$ *f)* $\begin{bmatrix} 5 & 1 & 3 \\ 4 & 1 & 1 \\ 4 & 5 & 1 \end{bmatrix}$ *g)* $\begin{bmatrix} 3 & 2 & 5 & 7 \\ 0 & 0 & 2 & 0 \\ 0 & 4 & 5 & 0 \\ 2 & 1 & 2 & 4 \end{bmatrix}$ *h)* $\begin{bmatrix} 0 & 2 & 1 & 0 \\ 1 & 2 & -3 & 4 \\ 5 & 6 & -7 & 8 \\ 1 & 2 & 3 & -2 \end{bmatrix}$

Exercise* A.6.102*: *For which x are the following matrices singular (not invertible).*

a) $\begin{bmatrix} 1 & 3 \\ 1 & x \end{bmatrix}$ *b)* $\begin{bmatrix} 3 & x \\ 1 & 3 \end{bmatrix}$ *c)* $\begin{bmatrix} x & 3 \\ 3 & x \end{bmatrix}$ *d)* $\begin{bmatrix} x & 1 & 0 \\ 1 & 4 & 0 \\ 1 & 6 & 2 \end{bmatrix}$

Exercise* A.6.103*: *Compute*

$$\det\left(\begin{bmatrix} 3 & 4 & 7 & 12 \\ 0 & -1 & 9 & -8 \\ 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 2 \end{bmatrix}^{-1}\right)$$

without computing the inverse.

Exercise* A.6.104** (challenging):*** *Find all the x that make the matrix inverse*

$$\begin{bmatrix} 1 & 2 \\ 1 & x \end{bmatrix}^{-1}$$

have only integer entries (no fractions). Note that there are two answers.

Appendix B

Complex numbers and Euler's formula

A polynomial may have complex roots. The equation $r^2 + 1 = 0$ has no real roots, but it does have two complex roots. Here we review some properties of complex numbers.

Complex numbers may seem a strange concept, especially because of the terminology. There is nothing imaginary or really complicated about complex numbers. A complex number is simply a pair of real numbers, (a, b). Think of a complex number as a point in the plane. We add complex numbers in the straightforward way: $(a, b) + (c, d) = (a + c, b + d)$. We define multiplication by

$$(a, b) \times (c, d) \overset{\text{def}}{=} (ac - bd, ad + bc).$$

It turns out that with this multiplication rule, all the standard properties of arithmetic hold. Further, and most importantly $(0, 1) \times (0, 1) = (-1, 0)$.

Generally we write (a, b) as $a + ib$, and we treat i as if it were an unknown. When b is zero, then $(a, 0)$ is just the number a. We do arithmetic with complex numbers just as we would with polynomials. The property we just mentioned becomes $i^2 = -1$. So whenever we see i^2, we replace it by -1. For example,

$$(2 + 3i)(4i) - 5i = (2 \times 4)i + (3 \times 4)i^2 - 5i = 8i + 12(-1) - 5i = -12 + 3i.$$

The numbers i and $-i$ are the two roots of $r^2 + 1 = 0$. Some engineers use the letter j instead of i for the square root of -1. We use the mathematicians' convention and use i.

Exercise* B.0.1*: *Make sure you understand (that you can justify) the following identities:*

a) $i^2 = -1$, $i^3 = -i$, $i^4 = 1$,

b) $\dfrac{1}{i} = -i$,

c) $(3 - 7i)(-2 - 9i) = \cdots = -69 - 13i$,

d) $(3-2i)(3+2i) = 3^2 - (2i)^2 = 3^2 + 2^2 = 13$,

e) $\frac{1}{3-2i} = \frac{1}{3-2i}\frac{3+2i}{3+2i} = \frac{3+2i}{13} = \frac{3}{13} + \frac{2}{13}i$.

We also define the exponential e^{a+ib} of a complex number. We do this by writing down the Taylor series and plugging in the complex number. Because most properties of the exponential can be proved by looking at the Taylor series, these properties still hold for the complex exponential. For example the very important property: $e^{x+y} = e^x e^y$. This means that $e^{a+ib} = e^a e^{ib}$. Hence if we can compute e^{ib}, we can compute e^{a+ib}. For e^{ib} we use the so-called *Euler's formula*.

Theorem B.0.1 (Euler's formula)**.**

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta \qquad \textit{and} \qquad e^{-i\theta} = \cos\theta - i\sin\theta.}$$

In other words, $e^{a+ib} = e^a\big(\cos(b) + i\sin(b)\big) = e^a\cos(b) + ie^a\sin(b)$.

Exercise* B.0.2*: *Using Euler's formula, check the identities:*

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \textit{and} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

Exercise* B.0.3*: *Double angle identities: Start with $e^{i(2\theta)} = \big(e^{i\theta}\big)^2$. Use Euler on each side and deduce:*

$$\cos(2\theta) = \cos^2\theta - \sin^2\theta \qquad \textit{and} \qquad \sin(2\theta) = 2\sin\theta\cos\theta.$$

For a complex number $a + ib$ we call a the *real part* and b the *imaginary part* of the number. Often the following notation is used,

$$\operatorname{Re}(a + ib) = a \qquad \text{and} \qquad \operatorname{Im}(a + ib) = b.$$

Further Reading

[BM] Paul W. Berg and James L. McGregor, *Elementary Partial Differential Equations*, Holden-Day, San Francisco, CA, 1966.

[BD] William E. Boyce and Richard C. DiPrima, *Elementary Differential Equations and Boundary Value Problems*, 11th edition, John Wiley & Sons Inc., New York, NY, 2017.

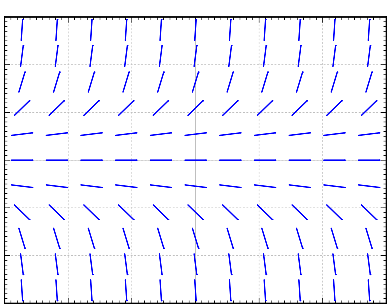
[EP] C.H. Edwards and D.E. Penney, *Differential Equations and Boundary Value Problems: Computing and Modeling*, 5th edition, Pearson, 2014.

[F] Stanley J. Farlow, *An Introduction to Differential Equations and Their Applications*, McGraw-Hill, Inc., Princeton, NJ, 1994. (Published also by Dover Publications, 2006.)

[I] E.L. Ince, *Ordinary Differential Equations*, Dover Publications, Inc., New York, NY, 1956.

[T] William F. Trench, *Elementary Differential Equations with Boundary Value Problems*. Books and Monographs. Book 9. 2013. https://digitalcommons.trinity.edu/mono/9

Solutions to Selected Exercises

0.2.101: Compute $x' = -2e^{-2t}$ and $x'' = 4e^{-2t}$. Then $(4e^{-2t}) + 4(-2e^{-2t}) + 4(e^{-2t}) = 0$.
0.2.102: Yes.
0.2.103: $y = x^r$ is a solution for $r = 0$ and $r = 2$.
0.2.104: $C_1 = 100$, $C_2 = -90$
0.2.105: $\varphi = -9e^{8s}$
0.2.106: a) $x = 9e^{-4t}$ b) $x = \cos(2t) + \sin(2t)$ c) $p = 4e^{3q}$ d) $T = 3\sinh(2x)$
0.3.101: a) PDE, equation, second order, linear, nonhomogeneous, constant coefficient.
b) ODE, equation, first order, linear, nonhomogeneous, not constant coefficient, not autonomous.
c) ODE, equation, seventh order, linear, homogeneous, constant coefficient, autonomous.
d) ODE, equation, second order, linear, nonhomogeneous, constant coefficient, autonomous.
e) ODE, system, second order, nonlinear.
f) PDE, equation, second order, nonlinear.
0.3.102: equation: $a(x)y = b(x)$, solution: $y = \frac{b(x)}{a(x)}$.
0.3.103: $k = 0$ or $k = 1$
1.1.101: $y = e^x + \frac{x^2}{2} + 9$
1.1.102: $x = (3t - 2)^{1/3}$
1.1.103: $x = \sin^{-1}\left(t + 1/\sqrt{2}\right)$
1.1.104: 170
1.1.105: If $n \neq 1$, then $y = \left((1 - n)x + 1\right)^{1/(1-n)}$. If $n = 1$, then $y = e^x$.
1.1.106: The equation is $r' = -C$ for some constant C. The snowball will be completely melted in 25 minutes from time $t = 0$.
1.1.107: $y = Ax^3 + Bx^2 + Cx + D$, so 4 constants.
1.2.101:  $y = 0$ is a solution such that $y(0) = 0$.

1.2.102: Yes a solution exists. The equation is $y' = f(x, y)$ where $f(x, y) = xy$. The function $f(x, y)$ is continuous and $\frac{\partial f}{\partial y} = x$, which is also continuous near $(0, 0)$. So a solution exists and is unique. (In fact, $y = 0$ is the solution.)

1.2.103: No, the equation is not defined at $(x, y) = (1, 0)$.

1.2.104: a) $y' = \cos y$, b) $y' = y\cos(x)$, c) $y' = \sin x$. Justification left to reader.

1.2.105: Picard does not apply as f is not continuous at $y = 0$. The equation does not have a continuously differentiable solution. Suppose it did. Notice that $y'(0) = 1$. By the first derivative test, $y(x) > 0$ for small positive x. But then for those x, we have $y'(x) = f\big(y(x)\big) = 0$. It is not possible for y' to be continous, $y'(0) = 1$ and $y'(x) = 0$ for arbitrarily small positive x.

1.2.106: The solution is $y(x) = \int_{x_0}^{x} f(s)\,ds + y_0$, and this does indeed exist for every x.

1.3.101: $y = Ce^{x^2}$

1.3.102: $x = e^{t^3} + 1$

1.3.103: $x^3 + x = t + 2$

1.3.104: $y = \frac{1}{1-\ln x}$

1.3.105: $\sin(y) = -\cos(x) + C$

1.3.106: The range is approximately 7.45 to 12.15 minutes.

1.3.107: a) $x = \frac{1000e^t}{e^t+24}$. b) 102 rabbits after one month, 861 after 5 months, 999 after 10 months, 1000 after 15 months.

1.4.101: $y = Ce^{-x^3} + 1/3$

1.4.102: $y = 2e^{\cos(2x)+1} + 1$

1.4.103: 250 grams

1.4.104: $P(5) = 1000e^{2\times 5-0.05\times 5^2} = 1000e^{8.75} \approx 6.31 \times 10^6$

1.4.105: $Ah' = I - kh$, where k is a constant with units m^2/s.

1.5.101: $y = \frac{2}{3x-2}$

1.5.102: $y = \frac{3-x^2}{2x}$

1.5.103: $y = \left(7e^{3x} + 3x + 1\right)^{1/3}$

1.5.104: $y = \pm\sqrt{x^2 - \ln(C - x)}$

1.6.101: a) 0, 1, 2 are critical points. b) $x = 0$ is unstable (semistable), $x = 1$ is stable, and $x = 2$ is unstable. c) 1

1.6.102: a) There are no critical points. b) ∞

1.6.103: a) $\frac{dx}{dt} = kx(M - x) + A$ b) $\frac{kM+\sqrt{(kM)^2+4Ak}}{2k}$

1.6.104: a) α is a stable critical point, β is an unstable one. b) α, c) α, d) ∞ or DNE.

1.7.101: Approximately: 1.0000, 1.2397, 1.3829

1.7.102: a) 0, 8, 12 b) $x(4) = 16$, so errors are: 16, 8, 4. c) Factors are 0.5, 0.5, 0.5.

1.7.103: a) 0, 0, 0 b) $x = 0$ is a solution so errors are: 0, 0, 0.

1.7.104: a) Improved Euler: $y(1) \approx 3.3897$ for $h = 1/4$, $y(1) \approx 3.4237$ for $h = 1/8$, b) Standard Euler: $y(1) \approx 2.8828$ for $h = 1/4$, $y(1) \approx 3.1316$ for $h = 1/8$, c) $y = 2e^x - x - 1$, so $y(2)$ is approximately 3.4366. d) Approximate errors for improved Euler: 0.046852 for $h = 1/4$, and 0.012881 for $h = 1/8$. For standard Euler: 0.55375 for $h = 1/4$, and 0.30499 for $h = 1/8$. Factor is approximately 0.27 for improved Euler, and 0.55 for standard Euler.

1.8.101: a) $e^{xy} + \sin(x) = C$ b) $x^2 + xy - 2y^2 = C$ c) $e^x + e^y = C$ d) $x^3 + 3xy + y^3 = C$

1.8.102: a) Integrating factor is y, equation becomes $dx + 3y^2\,dy = 0$. b) Integrating factor is e^x, equation becomes $e^x\,dx - e^{-y}\,dy = 0$. c) Integrating factor is y^2, equation becomes $(\cos(x) + y)\,dx + x\,dy = 0$. d) Integrating factor is x, equation becomes $(2xy + y^2)\,dx + (x^2 + 2xy)\,dy = 0$.

1.8.103: a) The equation is $-f(x)\,dx + \frac{1}{g(y)}\,dy = 0$, and this is exact because $M = -f(x)$, $N = \frac{1}{g(y)}$, so $M_y = 0 = N_x$. b) $-x\,dx + \frac{1}{y}\,dy = 0$, leads to potential function $F(x, y) = -\frac{x^2}{2} + \ln|y|$, solving $F(x, y) = C$ leads to the same solution as the example.

2.1.101: $y_1 = C_1 e^{3x}$, $y_2 = y(x) = C_2 e^x + \frac{C_1}{2} e^{3x}$, $y_3 = y(x) = C_3 e^x + \frac{C_1}{2} e^{3x}$

2.1.102: $x = \frac{5}{3} e^{2t} - \frac{2}{3} e^{-t}$, $y = \frac{5}{3} e^{2t} + \frac{4}{3} e^{-t}$

2.1.103: $x_1' = x_2$, $x_2' = x_3$, $x_3' = x_1 + t$

2.1.104: $y_3' + y_1 + y_2 = t$, $y_4' + y_1 - y_2 = t^2$, $y_1' = y_3$, $y_2' = y_4$

2.1.105: $x_1 = x_2 = at$. Explanation of the intuition is left to reader.

2.1.106: a) Left to reader. b) $x_1' = \frac{r-s}{V} x_2 - \frac{r}{V} x_1$, $x_2' = \frac{r}{V}(x_1 - x_2)$. c) As t goes to infinity, both x_1 and x_2 go to zero, explanation is left to reader.

2.2.101: -15

2.2.102: -2

2.2.103: $\vec{x} = \begin{bmatrix} 15 \\ -5 \end{bmatrix}$

2.2.104: a) $\begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix}$ b) $\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1/c \end{bmatrix}$

2.3.101: Yes.

2.3.102: No. $2 \begin{bmatrix} \cosh(t) \\ 1 \end{bmatrix} - \begin{bmatrix} e^t \\ 1 \end{bmatrix} - \begin{bmatrix} e^{-t} \\ 1 \end{bmatrix} = \vec{0}$

2.3.103: $\begin{bmatrix} x \\ y \end{bmatrix}' = \begin{bmatrix} 3 & -1 \\ t & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e^t \\ 0 \end{bmatrix}$

2.3.104: a) $\vec{x}\,' = \begin{bmatrix} 0 & 2t \\ 0 & 2t \end{bmatrix} \vec{x}$ b) $\vec{x} = \begin{bmatrix} C_2 e^{t^2} + C_1 \\ C_2 e^{t^2} \end{bmatrix}$

2.4.101: a) Eigenvalues: $4, 0, -1$ Eigenvectors: $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix}$
b) $\vec{x} = C_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} e^{4t} + C_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + C_3 \begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} e^{-t}$

2.4.102: a) Eigenvalues: $\frac{1+\sqrt{3}i}{2}, \frac{1-\sqrt{3}i}{2}$, Eigenvectors: $\begin{bmatrix} -2 \\ 1-\sqrt{3}i \end{bmatrix}, \begin{bmatrix} -2 \\ 1+\sqrt{3}i \end{bmatrix}$
b) $\vec{x} = C_1 e^{t/2} \begin{bmatrix} -2\cos\left(\frac{\sqrt{3}t}{2}\right) \\ \cos\left(\frac{\sqrt{3}t}{2}\right)+\sqrt{3}\sin\left(\frac{\sqrt{3}t}{2}\right) \end{bmatrix} + C_2 e^{t/2} \begin{bmatrix} -2\sin\left(\frac{\sqrt{3}t}{2}\right) \\ \sin\left(\frac{\sqrt{3}t}{2}\right)-\sqrt{3}\cos\left(\frac{\sqrt{3}t}{2}\right) \end{bmatrix}$

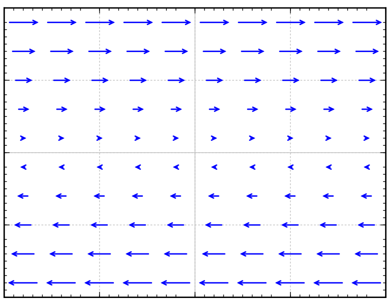
2.4.103: $\vec{x} = C_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{t} + C_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{-t}$

2.4.104: $\vec{x} = C_1 \begin{bmatrix} \cos(t) \\ -\sin(t) \end{bmatrix} + C_2 \begin{bmatrix} \sin(t) \\ \cos(t) \end{bmatrix}$

2.5.101: a) Two eigenvalues: $\pm\sqrt{2}$ so the behavior is a saddle. b) Two eigenvalues: 1 and 2, so the behavior is a source. c) Two eigenvalues: $\pm 2i$, so the behavior is a center (ellipses). d) Two eigenvalues: -1 and -2, so the behavior is a sink. e) Two eigenvalues: 5 and -3, so the behavior is a saddle.

2.5.102: Spiral source.

2.5.103:



The solution does not move anywhere if $y = 0$. When y is positive, the solution moves (with constant speed) in the positive x direction. When y is negative, the solution moves (with constant speed) in the negative x direction. It is not one of the behaviors we saw. Note that the matrix has a double eigenvalue 0 and the general solution is $x = C_1 t + C_2$ and $y = C_1$, which agrees with the description above.

2.6.101: Yes. To justify try to find a constant A such that $\sin(x) = Ae^{x}$ for all x.

2.6.102: No. $e^{x+2} = e^2 e^x$.

2.6.103: $y = 5$

2.6.104: $y = C_1 \ln(x) + C_2$

2.6.105: $y'' - 3y' + 2y = 0$

2.7.101: $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \bigl(a_1 \cos(\sqrt{3}\,t) + b_1 \sin(\sqrt{3}\,t)\bigr) + \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \bigl(a_2 \cos(\sqrt{2}\,t) + b_2 \sin(\sqrt{2}\,t)\bigr) +$ $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \bigl(a_3 \cos(t) + b_3 \sin(t)\bigr) + \begin{bmatrix} -1 \\ 1/2 \\ 2/3 \end{bmatrix} \cos(2t)$

2.7.102: $\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix} \vec{x}\,'' = \begin{bmatrix} -k & k & 0 \\ k & -2k & k \\ 0 & k & -k \end{bmatrix} \vec{x}$. Solution: $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \bigl(a_1 \cos(\sqrt{3k/m}\,t) + b_1 \sin(\sqrt{3k/m}\,t)\bigr)$ $+ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \bigl(a_2 \cos(\sqrt{k/m}\,t) + b_2 \sin(\sqrt{k/m}\,t)\bigr) + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \bigl(a_3 t + b_3\bigr)$.

2.7.103: $x_2 = (2/5)\cos(\sqrt{1/6}\,t) - (2/5)\cos(t)$

2.8.101: a) $3, 0, 0$ b) No defects. c) $\vec{x} = C_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} e^{3t} + C_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + C_3 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

2.8.102: a) $1, 1, 2$
b) Eigenvalue 1 has a defect of 1
c) $\vec{x} = C_1 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} e^t + C_2 \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right) e^t + C_3 \begin{bmatrix} 3 \\ 3 \\ -2 \end{bmatrix} e^{2t}$
2.8.103: a) $2, 2, 2$
b) Eigenvalue 2 has a defect of 2
c) $\vec{x} = C_1 \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} e^{2t} + C_2 \left(\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \right) e^{2t} + C_3 \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + \frac{t^2}{2} \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \right) e^{2t}$
2.8.104: $A = \begin{bmatrix} 5 & 5 \\ 0 & 5 \end{bmatrix}$
3.1.101: a) Critical points $(1, 0)$ and $(1, 1)$. At $(1, 0)$ using $u = x - 1$, $v = y$ the linearization is $u' = \pi v$, $v' = -v$. At $(1, 1)$ using $u = x - 1$, $v = y - 1$ the linearization is $u' = -\pi v$, $v' = v$.
b) Critical points $(0, 0)$ and $(0, -1)$. Using $u = x$, $v = y$ the linearization is $u' = u + v$, $v' = u$. At $(0, 0)$ using $u = x$, $v = y$ the linearization is $u' = u + v$, $v' = u$. At $(0, -1)$ using $u = x$, $v = y + 1$ the linearization is $u' = u - v$, $v' = v$.
c) Critical point $(1/2, -1/4)$. Using $u = x - 1/2$, $v = y + 1/4$ the linearization is $u' = -u + v$, $v' = u + v$.
3.1.102: 1) is c), 2) is a), 3) is b)
3.1.103: Critical points are $(0, 0, 0)$, and $(-1, 1, -1)$. The linearization at the origin using variables $u = x$, $v = y$, $w = z$ is $u' = u$, $v' = -v$, $z' = w$. The linearization at the point $(-1, 1, -1)$ using variables $u = x + 1$, $v = y - 1$, $w = z + 1$ is $u' = u - 2w$, $v' = -v - 2w$, $w' = w - 2u$.
3.1.104: $u' = f(u, v, w)$, $v' = g(u, v, w)$, $w' = 1$.
3.2.101: a) $(0, 0)$: saddle (unstable), $(1, 0)$: source (unstable), b) $(0, 0)$: spiral sink (asymptotically stable), $(0, 1)$: saddle (unstable), c) $(1, 0)$: saddle (unstable), $(0, 1)$: source (unstable)
3.2.102: a) $\frac{1}{2}y^2 + \frac{1}{3}x^3 - 4x = C$, critical points: $(-2, 0)$, an unstable saddle, and $(2, 0)$, a stable center. b) $\frac{1}{2}y^2 + e^x = C$, no critical points. c) $\frac{1}{2}y^2 + xe^x = C$, critical point at $(-1, 0)$ is a stable center.
3.2.103: Critical point at $(0, 0)$. Trajectories are $y = \pm\sqrt{2C - (1/2)x^4}$, for $C > 0$, these give closed curves around the origin, so the critical point is a stable center.
3.2.104: A critical point x_0 is stable if $f'(x_0) < 0$ and unstable when $f'(x_0) > 0$.
3.3.101: a) Critical points are $\omega = 0$, $\theta = k\pi$ for any integer k. When k is odd, we have a saddle point. When k is even we get a sink. b) The findings mean the pendulum will simply go to one of the sinks, for example $(0, 0)$ and it will not swing back and forth. The friction is too high for it to oscillate, just like an overdamped mass-spring system.
3.3.102: a) Solving for the critical points we get $(0, -h/d)$ and $(\frac{bh+ad}{ac}, \frac{a}{b})$. The Jacobian matrix at $(0, -h/d)$ is $\begin{bmatrix} a+bh/d & 0 \\ -ch/d & -d \end{bmatrix}$ whose eigenvalues are $a + bh/d$ and $-d$. The eigenvalues are real of opposite signs and we get a saddle. (In the application, however, we are only looking at the positive quadrant so this critical point is irrelevant.) At $(\frac{bh+ad}{ac}, \frac{a}{b})$ we get Jacobian matrix $\begin{bmatrix} 0 & -\frac{b(bh+ad)}{ac} \\ \frac{ac}{b} & \frac{bh+ad}{a} - d \end{bmatrix}$. b) For the specific numbers given, the second critical point

is $(\frac{550}{3}, 40)$ the matrix is $\begin{bmatrix} 0 & -11/6 \\ 3/25 & 1/4 \end{bmatrix}$, which has eigenvalues $\frac{5 \pm i\sqrt{327}}{40}$. Therefore there is a spiral source; the solution spirals outwards. The solution eventually hits one of the axes, $x = 0$ or $y = 0$, so something will die out in the forest.

3.3.103: The critical points are on the line $x = 0$. In the positive quadrant the y' is always positive and so the fox population always grows. The constant of motion is $C = y^a e^{-cx-by}$, for any C this curve must hit the y-axis (why?), so the trajectory will simply approach a point on the y axis somewhere and the number of hares will go to zero.

3.4.101: Use Bendixson–Dulac Theorem. a) $f_x + g_y = 1 + 1 > 0$, so no closed trajectories. b) $f_x + g_y = -\sin^2(y) + 0 < 0$ for all x, y except the lines given by $y = k\pi$ (where we get zero), so no closed trajectories. c) $f_x + g_y = y^2 + 0 > 0$ for all x, y except the line given by $y = 0$ (where we get zero), so no closed trajectories.

3.4.102: Using Poincaré–Bendixson Theorem, the system has a limit cycle, which is the unit circle centered at the origin, as $x = \cos(t) + e^{-t}$, $y = \sin(t) + e^{-t}$ gets closer and closer to the unit circle. Thus $x = \cos(t)$, $y = \sin(t)$ is the periodic solution.

3.4.103: $f(x, y) = y$, $g(x, y) = \mu(1-x^2)y - x$. So $f_x + g_y = \mu(1-x^2)$. The Bendixson–Dulac Theorem says there is no closed trajectory lying entirely in the set $x^2 < 1$.

3.4.104: The closed trajectories are those where $\sin(r) = 0$, therefore, all the circles centered at the origin with radius that is a multiple of π are closed trajectories.

3.5.101: Critical points: $(0, 0, 0)$, $(3\sqrt{8}, 3\sqrt{8}, 27)$, $(-3\sqrt{8}, -3\sqrt{8}, 27)$. Linearization at $(0, 0, 0)$ using $u = x$, $v = y$, $w = z$ is $u' = -10u + 10v$, $v' = 28u - v$, $w' = -(8/3)w$. Linearization at $(3\sqrt{8}, 3\sqrt{8}, 27)$ using $u = x - 3\sqrt{8}$, $v = y - 3\sqrt{8}$, $w = z - 27$ is $u' = -10u + 10v$, $v' = u - v - 3\sqrt{8}w$, $w' = 3\sqrt{8}u + 3\sqrt{8}v - (8/3)w$. Linearization at $(-3\sqrt{8}, -3\sqrt{8}, 27)$ using $u = x + 3\sqrt{8}$, $v = y + 3\sqrt{8}$, $w = z - 27$ is $u' = -10u + 10v$, $v' = u - v + 3\sqrt{8}w$, $w' = -3\sqrt{8}u - 3\sqrt{8}v - (8/3)w$.

A.1.101: a) $\sqrt{10}$ b) $\sqrt{14}$ c) 3

A.1.102: a) $\begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ b) $\begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$ c) $\left(\frac{2}{\sqrt{33}}, \frac{-5}{\sqrt{33}}, \frac{2}{\sqrt{33}}\right)$

A.1.103: a) $\begin{bmatrix} 9 \\ -2 \end{bmatrix}$ b) $\begin{bmatrix} -3 \\ 3 \end{bmatrix}$ c) $\begin{bmatrix} 5 \\ -3 \end{bmatrix}$ d) $\begin{bmatrix} -4 \\ 8 \end{bmatrix}$ e) $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$ f) $\begin{bmatrix} -8 \\ 3 \end{bmatrix}$

A.1.104: a) 20 b) 10 c) 20

A.1.105: a) $(3, -1)$ b) $(4, 0)$ c) $(-1, -1)$

A.2.101: a) $\begin{bmatrix} 7 & 4 & 4 \\ 2 & 3 & 4 \end{bmatrix}$ b) $\begin{bmatrix} 5 & -3 & 0 \\ 13 & 10 & 6 \\ -1 & 3 & 1 \end{bmatrix}$

A.2.102: a) $\begin{bmatrix} -1 & 13 \\ 9 & 14 \end{bmatrix}$ b) $\begin{bmatrix} 2 & -5 \\ 5 & 5 \end{bmatrix}$

A.2.103: a) $\begin{bmatrix} 22 & 31 \\ 42 & 44 \end{bmatrix}$ b) $\begin{bmatrix} 18 & 18 & 12 \\ 6 & 0 & 8 \\ 34 & 48 & -2 \end{bmatrix}$ c) $\begin{bmatrix} 11 & 12 & 36 & 14 \\ -2 & 4 & 5 & -2 \\ 13 & 38 & 20 & 28 \end{bmatrix}$ d) $\begin{bmatrix} -2 & -12 \\ 3 & 24 \\ 1 & 9 \end{bmatrix}$

A.2.104: a) $\begin{bmatrix} 1/2 \end{bmatrix}$ b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ c) $\begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$ d) $\begin{bmatrix} 1/2 & -1/4 \\ -1/2 & 1/2 \end{bmatrix}$

A.2.105: a) $\begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$ b) $\begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ c) $\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}$

A.3.101: a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1/3 \\ 0 & 0 & 0 \end{bmatrix}$ e) $\begin{bmatrix} 1 & 0 & 0 & 77/15 \\ 0 & 1 & 0 & -2/15 \\ 0 & 0 & 1 & -8/5 \end{bmatrix}$
f) $\begin{bmatrix} 1 & 0 & -1/2 & 0 \\ 0 & 1 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ g) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ h) $\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

A.3.102: a) $\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 5/2 & 1 & -3 \\ -1 & -1/2 & 3/2 \\ -1 & 0 & 1 \end{bmatrix}$

A.3.103: a) $x_1 = -2$, $x_2 = 7/3$ b) no solution c) $a = -3$, $b = 10$, $c = -8$ d) x_3 is free, $x_1 = -1 + 3x_3$, $x_2 = 2 - x_3$

A.3.104: a) $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ b) $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$

A.3.105: a) 3 b) 1 c) 2

A.3.106: a) $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 0 & 1/3 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & -1/3 \end{bmatrix}$

A.3.107: a) $\begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 7 \\ 6 \end{bmatrix}$, $\begin{bmatrix} 7 \\ 6 \\ 2 \end{bmatrix}$ b) $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ c) $\begin{bmatrix} 0 \\ 6 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 3 \\ 7 \end{bmatrix}$

A.3.108: $\begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}$

A.4.101: a) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ dimension 2, b) $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ dimension 2, c) $\begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 5 \\ -1 \\ 5 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 3 \\ -4 \end{bmatrix}$ dimension 3, d) $\begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$ dimension 2, e) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ dimension 1, f) $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$ dimension 2

A.4.102: a) $\begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 0 \\ 3 \\ -1 \end{bmatrix}$ b) $\begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$ c) $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ d) $\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

A.4.103: a) 3 b) 2 c) 3 d) 2 e) 3

A.5.101: $s = -2$
A.5.102: $\theta \approx 0.3876$
A.5.103: a) -15 b) -1 c) 28
A.5.104: a) $(-1/2, 0, \frac{1}{2})$ b) $(0, 0, 0)$ c) $(2, 0, -2)$
A.5.105: a) $(1, 1, -1) - (2, -1, 1) + 2(1, -5, 3)$ b) $2(2, -1, 1) + (1, -5, 3)$ c) $2(1, 1, -1) - 2(2, -1, 1) + 2(1, -5, 3)$
A.5.106: $(2, -1, 1)$, $(2/3, 8/3, 4/3)$
A.5.107: $(1, 1, -1)$, $(0, 1, 1)$, $(4/3, -2/3, 2/3)$
A.6.101: a) -2 b) 8 c) 0 d) -6 e) -3 f) 28 g) 16 h) -24
A.6.102: a) 3 b) 9 c) 3 d) $1/4$
A.6.103: 12
A.6.104: 1 and 3

Index

acceleration, 22
adding vectors, 161
addition of matrices, 79
algebraic multiplicity, 118
almost linear, 131
antiderivative, 20
antidifferentiate, 20
associated homogeneous equation, 90
associative law, 174
asymptotically stable, 132
asymptotically stable limit cycle, 148
attractor, 155
augmented matrix, 84, 182
autonomous, 16
autonomous equation, 49
autonomous system, 76

backsubstitution, 185
basis, 165, 195
Bendixson–Dulac Theorem, 149
Bernoulli equation, 45
Burger's equation, 15

cartesian plane, 160
catenary, 11
Cauchy–Euler equation, 107
center, 101, 133
chaotic systems, 152
Chebyshev's equation of order 1, 107
closed curves, 136
closed trajectory, 147
coefficients, 15
cofactor, 82, 212
cofactor expansion, 82, 211
column space, 190
column vector, 79, 161
commute, 81
complete eigenvalue, 118
complex conjugate, 95
complex number, 217
conservative equation, 135
conservative vector field, 62
consistent, 186
constant coefficient, 16, 89
constant of motion, 144
critical point, 49, 127

damped nonlinear pendulum equation, 145
defect, 119
defective eigenvalue, 119
deficient matrix, 119
dependent variable, 7
determinant, 81, 210
diagonal matrix, 109, 177
differential equation, 7
direction, 160, 163
direction field, 76
displacement vector, 109
distance, 22
distributive law, 174
dot product, 80, 172
Duffing equation, 153
dynamic damping, 117

echelon form, 183
eigenvalue, 92
eigenvector, 92
element of a matrix, 79, 164
elementary operations, 182
elementary row operations, 182
elimination, 181
ellipses (vector field), 101

entry of a matrix, 79, 164
equilibrium, 127
equilibrium solution, 49, 127
euclidean space, 160
Euler's equation, 107
Euler's formula, 218
Euler's method, 55
exact equation, 61
existence and uniqueness, 28, 105
existence and uniqueness for systems, 77
exponential growth, 14
exponential growth model, 9

first order differential equation, 7
first order linear equation, 38
first order linear system of ODEs, 88
first order method, 56
first order system, 71
forced motion
    systems, 115
four fundamental equations, 10
fourth order method, 58
Fredholm alternative, 199
free variable, 85, 186
fundamental matrix, 89
fundamental matrix solution, 89

Gauss–Jordan elimination, 183
Gaussian elimination, 183
general solution, 10
general solution to a system, 72
generalized eigenvectors, 119, 122
Genius software, 5
geometric multiplicity, 118
Gram–Schmidt process, 207

Hamiltonian, 135
harmonic conjugate, 69
harmonic function, 69
harvesting, 51
heat equation, 14
Hermite's equation of order 2, 107
homogeneous, 16
homogeneous equation, 46
homogeneous linear equation, 103
homogeneous matrix equation, 187
homogeneous system, 89
Hooke's law, 108
hyperbolic cosine, 11
hyperbolic sine, 11

identity matrix, 80, 174
imaginary part, 218
implicit solution, 33
Improved Euler's method, 60
inconsistent, 186
inconsistent system, 85
indefinite integral, 20
independent variable, 7
inhomogeneous, 16
initial condition, 10
initial condition for a system, 72
inner product, 80, 172, 202
integrate, 20
integrating factor, 38
integrating factor method, 38
invertible matrix, 81, 175
IODE software, 5
isolated critical point, 131

Jacobian matrix, 128

kernel, 187

leading entry, 85, 183
Leibniz notation, 21, 31
limit cycle, 147
linear combination, 104, 165, 187
linear equation, 15, 38, 103
linear first order system, 72
linear mapping, 164
linear operator, 104
linear systems, 72
linearization, 128
linearly dependent, 188
linearly independent, 105, 188
    for vector-valued functions, 89

logistic equation, 49
    with harvesting, 51
Lorenz attractor, 157
Lorenz system, 156
Lotka–Volterra, 142
lower triangular, 212

magnitude, 161
mass matrix, 109
mathematical model, 9
mathematical solution, 9
matrix, 79, 164
matrix inverse, 81, 175
matrix product, 173
matrix-valued function, 88
Maxwell's equations, 14
mechanical vibrations, 14
multiplication of complex numbers, 217
multiplicity of an eigenvalue, 118

n-dimensional space, 160
natural frequency, 111
natural mode of oscillation, 111
Newton's law of cooling, 14, 34, 42, 49
Newton's second law, 74, 108
nonhomogeneous, 16
nonlinear equation, 15
normal mode of oscillation, 111
nullity, 197
nullspace, 187

ODE, 8, 14
operator, 104, 164
order, 15
ordinary differential equation, 8
Ordinary differential equations, 14
orthogonal, 203
orthogonal basis, 205
orthogonal projection, 204
orthonormal basis, 205

parallelogram, 82, 210
partial differential equation, 8
Partial differential equations, 14
particular solution, 10
PDE, 8, 14
periodic orbit, 148
phase diagram, 50, 126
phase plane portrait, 76
phase portrait, 50, 76, 126
Picard's theorem, 28, 77
pivot, 183
Poincaré section, 154
Poincaré–Bendixson Theorem, 148
potential function, 61
predator-prey, 142
product of matrices, 80, 173
projection
    orthogonal, 204

rank, 189
real part, 218
real-world problem, 9
reduced row echelon form, 85, 183
reduction of order method, 106
relaxation oscillation, 147
resonance, 116
row echelon form, 183
row reduction, 183
row space, 190
row vector, 79, 165
Runge–Kutta method, 59

saddle point, 100
scalar, 79, 162
scalar multiplication, 79
scale a vector, 162
second order differential equation, 11
second order linear differential equation, 103
second order method, 57
second order system, 71
semistable critical point, 51
separable, 31
Sierpinski triangle, 158
simply connected region, 149
singular matrix, 81, 175
singular solution, 33

sink, 99
slope field, 25
solution, 7
solution curve, 76
solution to a system, 72
source, 99
span, 189
spiral sink, 102
spiral source, 101
square matrix, 165
stable center, 133
stable critical point, 49, 131
stable node, 99
standard basis vectors, 165
standard inner product, 202
stiff problem, 59
stiffness matrix, 109
strange attractor, 154
subspace, 195
subtracting vectors, 162
superposition, 89, 104
symmetric matrix, 178
system of differential equations, 14, 71

tedious, 152
three mass system, 108
total derivative, 61
trajectory, 76
transformation, 164
transport equation, 14
transpose, 80, 178

undamped motion
    systems, 108
undetermined coefficients
    for second order systems, 115
unit vector, 163
unstable critical point, 49, 131
unstable node, 99
upper triangular, 212
upper triangular matrix, 120

Van der Pol oscillator, 147

vector, 79, 160
vector field, 62, 76, 126
vector-valued function, 88, 163
velocity, 22

wave equation in 2 dimensions, 14