

Multivariable Mathematics for Data Science

John Tipton

Created 1 January 2020
Last updated 05 November, 2021

Contents

1 Preface	7
1.1 Getting started in R	7
1.2 Some videos that explain useful concepts of linear algebra	7
1.3 Notation	7
2 Linear Systems of Equations	15
2.1 Linear Systems of equations	16
2.2 Reduce row echelon form	35
3 Vectors spaces	41
3.1 Vectors	41
3.2 Vector addition	44
3.3 Span	54
4 Matrix equations	57
4.1 Solutions of matrix equations	57
4.2 Existence of solutions	58
4.3 Matrix-vector multiplication	58
4.4 Properties of matrix-vector multiplication	59
4.5 Solutions of linear systems	60
4.6 Solutions to nonhomogeneous systems	61
4.7 Finding solutions	65
5 Linear independence	67

6 Linear Transformations	69
6.1 Linear Transformations	72
6.2 Types of matrix transformations	73
6.3 Properties of matrix transformations	89
7 Matrix operations	93
7.1 ToDo	93
7.2 Properties of matrices	93
8 Matrix Inverses	101
8.1 Elementary matrices	104
8.2 Finding the inverse of \mathbf{A}	107
8.3 The Invertible Matrix Theorem	107
8.4 Invertible Linear Transformations	108
9 Block Matrices	109
9.1 Block Matrix Addition	111
9.2 Block Matrix Multiplication	111
9.3 The column-row matrix product	112
9.4 Special Block Matrices	112
10 Matrix Factorizations	115
10.1 The LU factorization	115
10.2 Obtaining the LU factorization	119
10.3 The Cholesky factor	119
11 Subspaces of \mathcal{R}^n	123
11.1 Special subspaces: column space and null space	125
11.2 The basis of a subspace	128
12 Dimension and Rank	137
12.1 Coordinate systems	137
12.2 Dimension of a subspace	139
12.3 Rank	142

CONTENTS	5
13 Determinants	145
13.1 Determinants of 2×2 matrices	147
13.2 Determinants of $n \times n$ matrices	147
13.3 Properties of determinants	150
13.4 Cramer's Rule and Determinants	151
14 Determinants and volumes	153
14.1 Volumes of Parallelpipeds	158
14.2 Volumes of Linear Transformations	158
15 Vector Spaces and Subspaces	159
15.1 Null space and column space	159
16 Linearly independent sets and bases	165
16.1 Bases for $\text{null}(\mathbf{A})$ and $\text{col}(\mathbf{A})$	166
17 Coordinate Systems and Dimension	169
17.1 Coordinates in \mathcal{R}^n	172
17.2 Dimension of a vector space	174
17.3 Subspaces of finite dimension	175
17.4 Dimensions of $\text{null}(\mathbf{A})$ and $\text{col}(\mathbf{A})$	175
18 Rank	177
18.1 Rank	177
19 Change of basis	179
20 Eigenvectors and Eigenvalues	181
20.1 Eigenspaces	182
21 The Characteristic Equation	187
21.1 Similarity	188
21.2 The geometric interpretation of similar matrices	190
22 Diagonalization	191

23 Inner product, length, and orthogonality	195
23.1 Distance	197
23.2 Orthogonal vectors	197
23.3 Angles between vectors	197
23.4 Orthogonal sets	198
23.5 Orthogonal projections	199
24 Graphs and Limits	205
24.1 Graphs and level curves	205
24.2 Limits	211
24.3 Continuity	214
25 Partial Derivatives	215
25.1 Higher-order partial derivatives	217
26 The chain rule	219
26.1 The chain rule with one independent variable	219
26.2 The chain rule with several independent variables	220
26.3 The chain rule in matrix notation	220
27 The gradient and directional derivatives	223
27.1 The Gradient	226
28 Tangent planes and linear approximations	229
29 Minimums and Maximums	235
29.1 Local minimums	235
29.2 Global maximums and minimums	241
29.3 Grid search optimization	241
29.4 Gradient Descent	242

Chapter 1

Preface

This book will introduce students to multivariable Calculus and linear algebra methods and techniques to be successful in data science, statistics, computer science, and other data-driven, computational disciplines.

The motivation for this text is to provide both a theoretical understanding of important multivariable methods used in data science as well as giving a hands-on experience using software. Throughout this text, we assume the reader has a solid foundation in univariate calculus (typically two semesters) as well as familiarity with a scripting language (e.g., R or python).

1.1 Getting started in R

TBD

1.2 Some videos that explain useful concepts of linear algebra

- 3 Blue 1 Brown – Essence of Linear Algebra
- 3 Blue 1 Brown – Vectors

1.3 Notation

For notation, we let lowercase Roman letters represent scalar numbers (e.g., $n = 5$, $d = 7$), lowercase bold letters represent vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where the elements x_1, \dots, x_n are scalars written in lowercase Roman. Note that vectors are assumed to follow a vertical notation where the elements of the vector (the x_i s are stacked on top of one another) and the order matters. For example, the vector

$$\mathbf{x} = \begin{pmatrix} 5 \\ 2 \\ 8 \end{pmatrix}$$

has the first element $x_1 = 5$, second element $x_2 = 2$ and third element $x_3 = 8$.

Note that the vector $\begin{pmatrix} 5 \\ 2 \\ 8 \end{pmatrix}$ is not the same as the vector $\begin{pmatrix} 8 \\ 2 \\ 5 \end{pmatrix}$ because the order of the elements matters.

We can also write the vector as

$$\mathbf{x} = (x_1, x_2, \dots, x_n)',$$

where the ' $'$ symbol represents the transpose function. For our example matrix,

we have $\begin{pmatrix} 5 \\ 2 \\ 8 \end{pmatrix}' = (5 \ 2 \ 8)$ which is the original vector but arranged in a row rather than a column. Likewise, the transpose of a row vector $(5 \ 2 \ 8)' = \begin{pmatrix} 5 \\ 2 \\ 8 \end{pmatrix}$ is a column vector. If \mathbf{x} is a column vector, we say that \mathbf{x}' is a row vector and if \mathbf{x} is a row vector, the \mathbf{x}' is a column vector.

To create a vector we can use the concatenate function `c()`. For example, the vector $\mathbf{x} = \begin{pmatrix} 5 \\ 2 \\ 8 \end{pmatrix}$ can be created as the R object using

```
x <- c(5, 2, 8)
```

where the `<-` assigns the values in the vector `c(5, 2, 8)` to the object named `x`. To print the values of `x`, we can use

```
x
## [1] 5 2 8
```

which prints the elements of `x`. Notice that R prints the elements of `x` in a row; however, `x` is a column vector. This inconsistency is present to allow the output to be printed in a manner easier to read (more numbers fit on a row). If we put the column vector into a `data.frame`, then the vector will be presented as a column vector

```
data.frame(x)
```

```
##   x
## 1 5
## 2 2
## 3 8
```

One can use the index operator `[]` to select specific elements of the vector `x`. For example, the first element of `x`, x_1 , is

```
x[1]
```

```
## [1] 5
```

and the third element of `x`, x_3 , is

```
x[3]
```

```
## [1] 8
```

The transpose function `t()` turns a column vector into a row vector (or a row vector into a column vector). For example the transpose \mathbf{x}' of `x` is

```
tx <- t(x)
tx
```

```
##      [,1] [,2] [,3]
## [1,]     5     2     8
```

where `tx` is R object storing the transpose of `x` and is a row vector. The transpose of `tx`. Notice the indices on the output of the row vector `tx`. The index operator `[1,]` selects the first row to `tx` and the index operator `[, 1]` gives the first column `tx`. Taking the transpose again gives us back the original column vector

```
t(tx)
```

```
##      [,1]
## [1,]    5
## [2,]    2
## [3,]    8
```

1.3.1 Matrices

We let uppercase bold letters **A**, **B**, etc., represent matrices. We define the matrix **A** with m rows and n columns as

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

with a_{ij} being the value of the matrix **A** in the i th row and the j th column.

If the matrix

$$\mathbf{A} = \begin{pmatrix} 5 & 7 & 1 \\ 5 & -22 & 2 \\ -14 & 5 & 99 \\ 42 & -3 & 0 \end{pmatrix},$$

the elements $a_{11} = 5$, $a_{12} = 7$, $a_{21} = 5$, and $a_{33} = 99$, etc.

In R, we can define the matrix A using the `matrix()` function

```
A <- matrix(
  data = c(5, 5, -14, 42, 7, -22, 5, -3, 1, 2, 99, 0),
  nrow = 4,
  ncol = 3
)

A

##      [,1] [,2] [,3]
## [1,]    5    7    1
## [2,]    5   -22    2
## [3,]   -14     5   99
## [4,]    42   -3    0
```

Notice in the above creation of **A**, we wrote defined the elements of the **A** using the columns stacked on top of one another. If we want to fill in the elements of **A** using the rows, we can add the option `byrow = TRUE` to the `matrix()` function

```
A <- matrix(
  data = c(5, 7, 1, 5, -22, 2, -14, 5, 99, 42, -3, 0),
  nrow = 4,
  ncol = 3,
  byrow = TRUE
)
A
```

```
##      [,1] [,2] [,3]
## [1,]    5    7    1
## [2,]   -22    2   -14
## [3,]    5    99    42
## [4,]   -3    0    42
```

To select the ij th elements of **A**, we use the subset operator `[` to select the element. For example, to get the element $a_{11} = 5$ in the first row and first column of **A**, we use

```
A[1, 1]
```

```
## [1] 5
```

The element $a_{3,3} = 99$ in the third row and third column can be selected using

```
A[3, 3]
```

```
## [1] 99
```

The matrix **A** can also be represented as a set of either column vectors $\{\mathbf{c}_j\}_{j=1}^n$ or row vectors $\{\mathbf{r}_i\}_{i=1}^m$. For example, the column vector representation is

$$\mathbf{A} = (\mathbf{c}_1 | \mathbf{c}_2 | \cdots | \mathbf{c}_n),$$

where the notation `|` is used to separate the vectors

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2m} \end{pmatrix}, \quad \dots, \quad \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

In R you can extract the columns using the [selection operator

```
c1 <- A[, 1] # first column
c2 <- A[, 2] # second column
c3 <- A[, 3] # third column
```

and you can give the column representation of the matrix **A** with with column bind function **cbind()**

```
cbind(c1, c2, c3)
```

```
##      c1  c2  c3
## [1,]  5   7   1
## [2,]  5  -22  2
## [3,] -14   5  99
## [4,]  42  -3   0
```

The row vector representation of **A** is

$$\mathbf{A} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{pmatrix},$$

where the row vectors \mathbf{r}_i are

$$\begin{aligned} \mathbf{r}_1 &= (a_{11}, a_{12}, \dots, a_{1n}) \\ \mathbf{r}_2 &= (a_{21}, a_{22}, \dots, a_{2n}) \\ &\vdots \\ \mathbf{r}_m &= (a_{m1}, a_{m2}, \dots, a_{mn}) \end{aligned}$$

In R you can extract the rows using the [selection operator

```
r1 <- A[1, ] # first row
r2 <- A[2, ] # second row
r3 <- A[3, ] # third row
r4 <- A[4, ] # fourth row
```

and you can give the row representation of the matrix **A** with with row bind function **rbind()**

```
rbind(r1, r2, r3, r4)
```

```
##      [,1] [,2] [,3]
## r1      5     7     1
## r2      5    -22     2
## r3    -14     5    99
## r4     42    -3     0
```


Chapter 2

Linear Systems of Equations

```
library(tidyverse)

# For 3-d plotting
# if devtools package not installed, install the package
if (!require(devtools)) {
  install.packages("devtools")
}
# if gg3D package not installed, install the package
if (!require(gg3D)) {
  devtools::install_github("AckerDWM/gg3D")
  library(gg3D)
}

## Warning in fun(libname, pkgname): no display name and no $DISPLAY environment
## variable

# if dasc2594 package not installed, install the package
if (!require(dasc2594)) {
  devtools::install_github("jtipton25/dasc2594")
  library(dasc2594)
}
```

2.1 Linear Systems of equations

2.1.1 Linear equations

Definition 2.1 (Linear Equations). Let x_1, x_2, \dots, x_n be **variables** with **coefficients** a_1, a_2, \dots, a_n , and b are fixed and known numbers. Then, we say

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

is a **linear equation**.

Example 2.1. The equation for a line with slope m and y -intercept b is

$$y = mx + b,$$

is a linear equation because it can be re-written as

$$y - mx = b,$$

where $a_1 = 1$, $a_2 = m$, $x_1 = y$ and $x_2 = x$.

Example 2.2. The equations

$$\sqrt{19}x_1 = (4 + \sqrt{2})x_2 - x_3 - 9 \quad \text{and} \quad -4x_1 + 5x_2 - 11 = x_3$$

are both linear equations because they can be written as

$$\sqrt{19}x_1 - (4 + \sqrt{2})x_2 + x_3 = -9 \quad \text{and} \quad -4x_1 + 5x_2 - x_3 = 11,$$

respectively.

Example 2.3. The equations

$$x_1 = x_2^2 + 3 \quad \text{and} \quad x_1 + x_2 - x_1x_2 = 16$$

are not linear equations because they do not meet the form of (2.1) (The first equation above has a quadratic power of x_2 and the second equation has a product of x_1 and x_2).

2.1.2 Systems of linear equations

Definition 2.2 (System of equations). A set of two or more linear equations that each contain the same set of variables is called a **system of linear equations**.

Example 2.4. The equations

$$\begin{aligned}x_1 + 4x_2 - x_3 &= 11 \\4x_1 + 5x_2 + 2x_3 &= 9.\end{aligned}$$

are a system of equations. Note that in the second equation, the coefficient for x_3 is 0, meaning we could re-write the above example as

$$\begin{aligned}x_1 + 4x_2 - x_3 &= 11 \\4x_1 + 5x_2 + 0x_3 &= 9.\end{aligned}$$

For the following, are these linear equations?

Example 2.5. $x_1 + 3x_1x_2 = 5$

Solution. This is a linear equation because the equation can be written as $a_1x_1 + a_2x_2 = b$ where $a_1 = 1$, $a_2 = 3$ and $b = 5$

Example 2.6. $5x + 7y + 8z = 11.2$

Solution. This is a linear equation because the equation can be written as $a_1x_1 + a_2x_2 + a_3x_3 = b$ where $a_1 = 5$, $a_2 = 7$, $a_3 = 8$ and $b = 11.2$. The variables $x_1 = x$, $x_2 = y$, and $x_3 = z$.

Example 2.7. $\frac{1}{4}y + \sqrt{2}z = 2^6$

Solution. This is a linear equation because the equation can be written as $a_1x_1 + a_2x_2 = b$ where $a_1 = \frac{1}{4}$, $a_2 = \sqrt{2}$, and $b = 2^6$. The variables $x_1 = y$, $x_2 = z$ enter the equation linearly.

Example 2.8. $x + 4y^2 = 9$

Solution. This is not a linear equation because the equation cannot be written as $a_1x_1 + a_2x_2 = b$ because $x_1 = x$ and $x_2 = y$ so that the equation is written as $a_1x_1 + a_2x_2^2 = b$ where $a_1 = 1$, $a_2 = 4$, and $b = 9$. The variable $x_2 = y$ enters the equation in a non-linear (quadratic) manner.

2.1.3 Solutions of linear systems

A fundamental question when presented with a linear system of equations is whether the system has a solution.

Definition 2.3 (Solution of systems of equations). A solution to a system means that there are numbers (s_1, s_2, \dots, s_n) that each of the variables x_1, x_2, \dots, x_n take that allow for all the equations to simultaneously be true.

Example 2.9. Is $\mathbf{x} = \begin{pmatrix} 5 \\ 3 \\ 6 \end{pmatrix}$ a solution to the system of equations

$$\begin{aligned} x_1 + 4x_2 - x_3 &= 11 \\ 4x_1 + 5x_2 + 0x_3 &= 9. \end{aligned}$$

Solution. First, we plug in the values for $x_1 = 5$, $x_2 = 3$ and $x_3 = 6$. In the first equation, we get

$$5 + 4 \times 3 - 6 = 11$$

which is true because $11 = 11$. Now we consider the second equation

$$4 \times 5 + 5 \times 3 + 0 \times 6 = 9.$$

which is false because $35 \neq 9$. Thus, $\mathbf{x} = \begin{pmatrix} 5 \\ 3 \\ 6 \end{pmatrix}$ is not a solution to the system of equations in the given example.

Example 2.10. Consider the system of equations

$$x + 4y = 8$$

$$4x + 5y = 7.$$

Solution. To find if a solution to this equation exists, we can do some algebra and take 4 times the top equation and then subtract the bottom equation, replacing the bottom equation with this new sum like

$$\begin{aligned} x + &\quad 4y = && 8 \\ 4x - 4 * (x) + 5y - 4 * (4y) &= 7 - 4 * (8), \end{aligned}$$

where the part of the equations in (·) is the top equation. This system of equations now simplifies to

$$\begin{aligned} x + 4y &= 8 \\ 0 + -11y &= -25, \end{aligned}$$

which gives $y = \frac{25}{11}$. Plugging this value into the top equation gives

$$\begin{aligned} x + 4\frac{25}{11} &= 8 \\ 0 + y &= \frac{25}{11}, \end{aligned}$$

where we can solve $x = 8 - \frac{100}{11} = -\frac{12}{11}$ giving the solution of the form

$$\begin{aligned} x + 0 &= -\frac{12}{11} \\ 0 + y &= \frac{25}{11}, \end{aligned}$$

In this case, the system of equation has the solution $x = -\frac{12}{11}$ and $y = \frac{25}{11}$. While finding the solution can be done algebraically, what does this mean visually (geometrically)? The original equations were

$$\begin{aligned} x + 4y &= 8 \\ 4x + 5y &= 7, \end{aligned}$$

which, writing y as a function of x define two lines:

$$\begin{aligned} 1) \quad y &= -\frac{x}{4} + 2 \\ 2) \quad y &= -\frac{4x}{5} + \frac{7}{5} \end{aligned}$$

Let's plot these equations in R and see what they look like

```
# define some grid points to evaluate the line
x <- seq(-2, 2, length = 1000)
dat <- data.frame(
  x = c(x, x),
  y = c(-x / 4 + 2, -4 / 5 * x + 7/5),
  equation = factor(rep(c(1, 2), each = 1000)))
)
glimpse(dat)
```

```

## Rows: 2,000
## Columns: 3
## $ x      <dbl> -2.000000, -1.995996, -1.991992, -1.987988, -1.983984, -1.979-
## $ y      <dbl> 2.500000, 2.498999, 2.497998, 2.496997, 2.495996, 2.494995, 2-
## $ equation <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
```

```

dat %>%
  ggplot(aes(x = x, y = y, color = equation, group = equation)) +
  geom_line() +
  scale_color_viridis_d(end = 0.8) +
  # solution x = -12/11, y = 25/11
  geom_point(aes(x = -12/11, y = 25/11), color = "red", size = 2) +
  ggtitle("Linear system of equations")

```

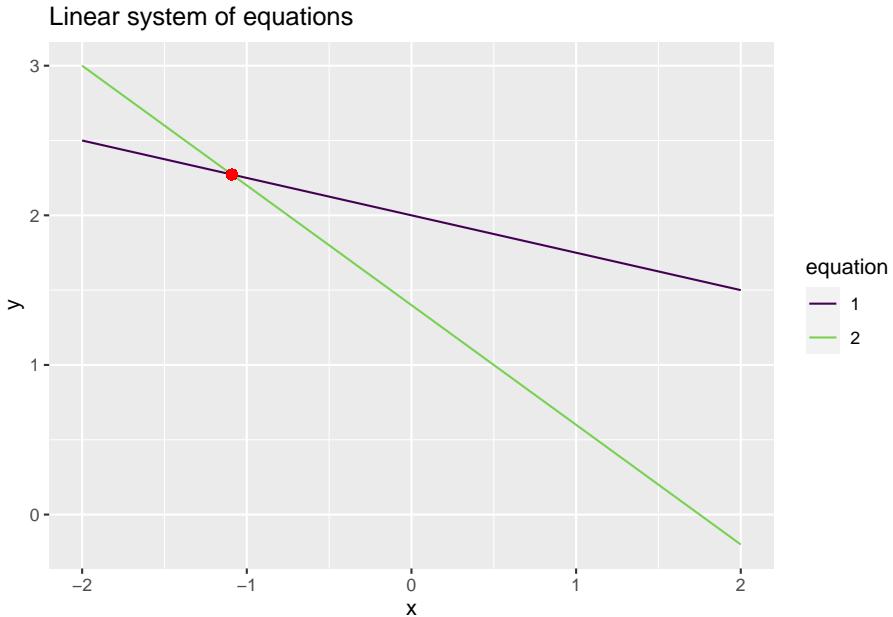


Figure 2.1: Linear system of equations with one solution

From this plot, it is clear that the solution to the system of equations is the location where the two lines intersect!

2.1.4 Types of solutions

Typically, there are 3 cases for the solutions to a system of linear equations

- 1) There are no solutions
 - 2) There is one solution (Figure 2.1)
 - 3) There are infinitely many solutions

Definition 2.4. A linear system of equations is called **consistent** if the system has either one or infinitely many solutions and is called **inconsistent** if the system has no solution.

2.1.4.1 There are no solutions:

Consider the system of linear equations

$$\begin{aligned} x + 4y &= 8 \\ 4x + 16y &\equiv 18. \end{aligned}$$

In this case, the linear equations are parallel lines and will never intersect so therefore there is no solution.

Solution. To find if a solution to this equation exists, we can do some algebra and take 4 times the top equation and then subtract the bottom equation, replacing the bottom equation with this new sum like

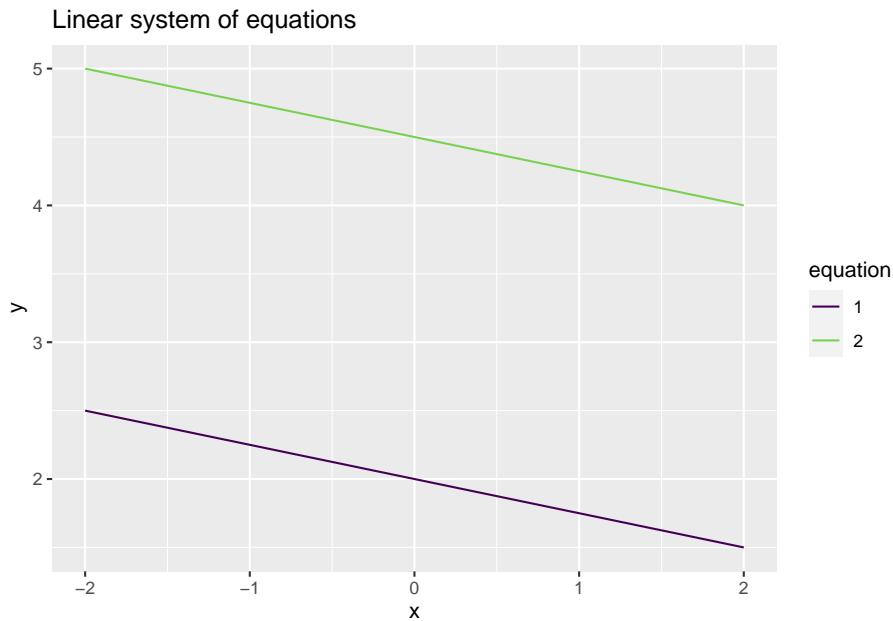


Figure 2.2: Linear system of equations with no solution

$$\begin{aligned} x + & \quad 4y = & 8 \\ 4x - 4(x) + 16y - 4 \times (4y) = & 18 - 4 \times (8). \end{aligned}$$

where the part of the equations in () is the top equation. This system of equations now simplifies to

$$\begin{aligned} x + 4y &= 8 \\ 0x + 0y &= -14, \end{aligned}$$

which gives $0 = -14$. From this, we see that we have reached a contradiction so there is not a solution to the system of equations.

2.1.4.2 There is one solution:

We have seen this example in Figure 2.1.

2.1.4.3 There are infinitely many solutions:

Consider the system of linear equations

$$\begin{aligned} x + 4y &= 8 \\ 4x + 16y &= 32. \end{aligned}$$

```
# define some grid points to evaluate the line
x <- seq(-2, 2, length = 1000)
dat <- data.frame(
  x = c(x, x),
  y = c(-x / 4 + 8 / 4, -4 * x / 16 + 32 / 16),
  equation = factor(rep(c(1, 2), each = 1000))
)
glimpse(dat)

## Rows: 2,000
## Columns: 3
## $ x      <dbl> -2.000000, -1.995996, -1.991992, -1.987988, -1.983984, -1.979~ 
## $ y      <dbl> 2.500000, 2.498999, 2.497998, 2.496997, 2.495996, 2.494995, 2~ 
## $ equation <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~ 

dat %>%
  ggplot(aes(x = x, y = y, color = equation, group = equation)) +
  geom_line() +
  scale_color_viridis_d(end = 0.8) +
  # solution x = -12/11, y = 25/11
  ggtitle("Linear system of equations")
```

In this case, the linear equations are perfectly overlapping lines and always intersect so therefore there are infinitely many solutions (all points on the line).

Definition 2.5. Two linear systems of equations are called **equivalent** if both systems share the same **solution set**.

For example, the system of equations

$$\begin{aligned} x_1 + 4x_2 - x_3 &= 11 \\ 4x_1 + 5x_2 + 2x_3 &= 9 \end{aligned}$$

and the system of equations

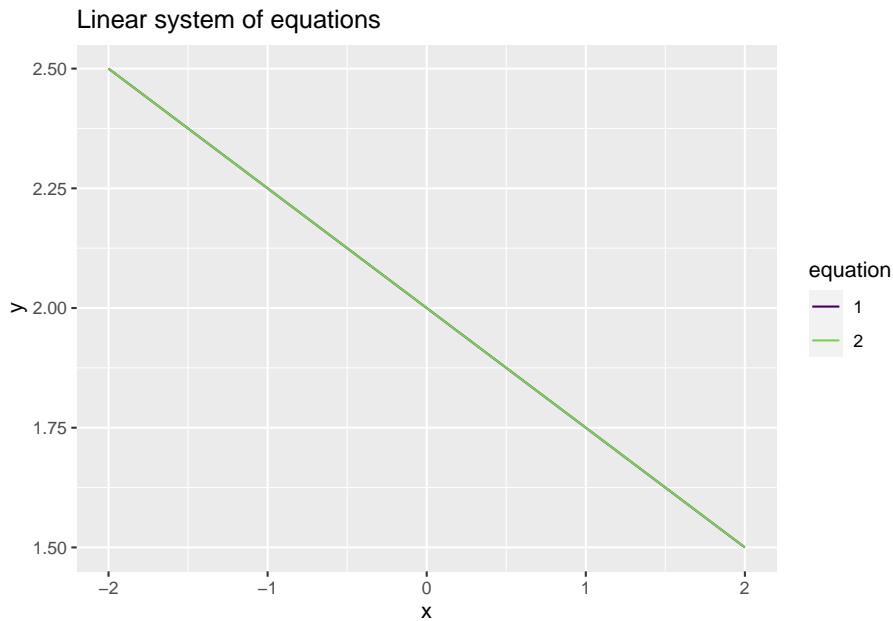


Figure 2.3: Linear system of equations with no solution

$$\begin{aligned} 2x_1 + 8x_2 - 2x_3 &= 22 \\ 8x_1 + 10x_2 + 4x_3 &= 18. \end{aligned}$$

have the same solution set (the second set of equations is just 2 times the first set of equations).

Example 2.11. For the following system of equations, determine if a solution(s) exist and if so, solve for the solution

$$\begin{aligned} 4x_1 + 5x_2 &= 8 \\ 9x_1 - 3x_2 &= 4. \end{aligned}$$

Solution. Multiply the first equation by $\frac{1}{4}$ and the second equation by $\frac{1}{9}$ to get

$$\begin{aligned} x_1 + \frac{5}{4}x_2 &= 2 \\ x_1 - \frac{1}{3}x_2 &= \frac{4}{9}. \end{aligned}$$

Subtract the first equation from the second equation

$$\begin{aligned}x_1 + \frac{5}{4}x_2 &= 2 \\0 - \left(-\frac{1}{3} - \frac{5}{4}\right)x_2 &= \frac{4}{9} - 2,\end{aligned}$$

which reduces to

$$\begin{aligned}x_1 + \frac{5}{4}x_2 &= 2 \\0 - \frac{19}{12}x_2 &= -\frac{14}{9},\end{aligned}$$

so that, dividing both sides of the second equation by $-\frac{19}{12}$ gives $x_2 = \frac{56}{57}$. Plugging this value of x_2 into the first equation gives

$$\begin{aligned}x_1 + \frac{5}{4} \left(\frac{56}{57}\right) &= 2 \\0 - x_2 &= \frac{56}{57}\end{aligned}$$

and subtracting $\frac{5}{4} \frac{56}{57}$ from both sides of the first equation gives $x_1 = \frac{44}{57}$. You can check these solutions by verifying that the following two equations hold

$$\begin{aligned}4 \left(\frac{44}{57}\right) + 5 \left(\frac{56}{57}\right) &= 8 \\9 \left(\frac{44}{57}\right) - 3 \left(\frac{56}{57}\right) &= 4\end{aligned}$$

which can be done in R using

```
x1 <- 44/57
x2 <- 56/57
all.equal(4 * x1 + 5 * x2, 8)
```

```
## [1] TRUE
```

```
all.equal(9 * x1 - 3 * x2, 4)
```

```
## [1] TRUE
```

Example 2.12. For the following system of equations, determine if a solution(s) exist and if so, solve for the solution

$$\begin{aligned} 7x_1 + 3x_2 + 4x_3 &= 5 \\ 4x_1 - 5x_2 &= -2 \end{aligned}$$

Solution. Multiply the first equation by $\frac{1}{7}$ and the second equation by $\frac{1}{4}$ to get

$$\begin{aligned} x_1 + \frac{3}{7}x_2 + \frac{4}{7}x_3 &= \frac{5}{7} \\ x_1 - \frac{5}{4}x_2 + 0x_3 &= \frac{-1}{2} \end{aligned}$$

Subtract the first equation from the second equation

$$\begin{aligned} x_1 + \frac{3}{7}x_2 + \frac{4}{7}x_3 &= \frac{5}{7} \\ 0x_1 + \left(-\frac{5}{4} - \frac{3}{7}\right)x_2 + -\frac{4}{7}x_3 &= -\frac{1}{2} - \frac{5}{7} \end{aligned}$$

which reduces to

$$\begin{aligned} x_1 + \frac{3}{7}x_2 + \frac{4}{7}x_3 &= \frac{5}{7} \\ 0 - \frac{47}{28}x_2 - \frac{4}{7}x_3 &= -\frac{17}{14}. \end{aligned}$$

Next, divide the second equation by $\frac{-47}{28}$ gives

$$\begin{aligned} x_1 + \frac{3}{7}x_2 + \frac{4}{7}x_3 &= \frac{5}{7} \\ 0 + x_2 + \frac{16}{47}x_3 &= \frac{34}{47}. \end{aligned}$$

Then, take the first equation and subtract $-\frac{3}{7}$ times the second row to get

$$\begin{aligned} x_1 + 0x_2 + \frac{20}{47}x_3 &= \frac{19}{47} \\ 0 + x_2 + \frac{16}{47}x_3 &= \frac{34}{47}. \end{aligned}$$

Which gives the solution $x_1 + \frac{20}{47}x_3 = \frac{19}{47}$, $x_2 + \frac{16}{47}x_3 = \frac{34}{47}$, and $x_3 = x_3$ is a free variable.

The solution can be checked in R using

```
# this is a free variable, you can assign it any value
x3 <- 3
x1 <- 19/47 - 20/47 * x3
x2 <- 34/47 - 16/47 * x3
# check the first equation
all.equal(7 * x1 + 3 * x2 + 4 * x3, 5)

## [1] TRUE

# check the second equation
all.equal(4 * x1 - 5 * x2 + 0 * x3, -2)
```

```
## [1] TRUE
```

Example 2.13. For the following system of equations, determine if a solution(s) exist and if so, solve for the solution

$$\begin{aligned} 4x_1 - 2x_2 &= 8 \\ 2x_1 + x_2 &= 7 \\ -3x_1 + 6x_2 &= 11 \end{aligned}$$

Solution. Multiply the first equation by $\frac{1}{4}$, the second equation by $\frac{1}{2}$, and the third equation by $-\frac{1}{3}$ to get

$$\begin{aligned} x_1 - \frac{1}{2}x_2 &= 2 \\ x_1 + \frac{1}{2}x_2 &= \frac{7}{2} \\ x_1 - 2x_2 &= -\frac{11}{3} \end{aligned}$$

Subtract the first equation from the second equation and also subtract the first equation from the third equation. Thus, we get

$$\begin{aligned} x_1 - \frac{1}{2}x_2 &= 2 \\ 0x_1 + x_2 &= \frac{3}{2} \\ 0x_1 - \frac{3}{2}x_2 &= -\frac{17}{3}. \end{aligned}$$

Then, multiply the third equation by $-\frac{2}{3}$ to get

$$\begin{aligned}x_1 - \frac{1}{2}x_2 &= 2 \\0x_1 + x_2 &= 3 \\0x_1 + x_2 &= -\frac{34}{9}.\end{aligned}$$

Notice that the second equation gives $x_2 = 3$ while the third equation gives $x_2 = -\frac{34}{9}$. This is a contradiction ($3 \neq -\frac{34}{9}$) which implies that the system of equations does not have a solution. Thus we say the system of equations is inconsistent.

2.1.5 Elementary row and column operations on matrices

The elementary row (column) operations include

- 1) swaps: swapping two rows (columns),
- 2) sums: replacing a row (column) by the sum itself and a multiple of another row (column)
- 3) scalar multiplication: replacing a row (column) by a scalar multiple times itself

Note that these operations are exactly what we used to solve the equation using algebra above (except for swapping rows).

Example 2.14. For the elementary row operations listed above, we demonstrate these using the matrix

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

The matrix **A** can be represented in R using

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
```

- 1) Swap the first and second rows.
- 2) Add -3 times the first row to the third row.
- 3) Multiply the second row by $\frac{1}{2}$.

Solution. Here we present examples of the elementary row operations

- 1) Swap the first and second rows.

The matrix with the first and second rows swapped is

$$\begin{pmatrix} 2 & 5 & 8 \\ 1 & 4 & 7 \end{pmatrix}$$

Notice that the first and second rows have now switched places. This can be done in R in a variety of ways. First, we extract the rows and place them “by hand” in R.

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
# extract the rows
row1 <- A[1, ]
row2 <- A[2, ]
# swap the rows
A[1, ] <- row2
A[2, ] <- row1
A

##      [,1] [,2] [,3]
## [1,]     2     5     8
## [2,]     1     4     7
## [3,]     3     6     9
```

Another way to do this is to use the function `rbind()` that binds rows together into a matrix.

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
# bind row 2, row 1, and row 3 together
rbind(A[2, ], A[1, ], A[3, ])

##      [,1] [,2] [,3]
## [1,]     2     5     8
## [2,]     1     4     7
## [3,]     3     6     9
```

Yet another way to swap the first two rows is to use a vectorized operation. This allows for fast and efficient coding and computing. The vectorized row swap is

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
# Take the 2nd, 1st, and 3rd row of A, in that order
A[c(2, 1, 3), ]
```

```
##      [,1] [,2] [,3]
## [1,]    2    5    8
## [2,]    1    4    7
## [3,]    3    6    9
```

- 2) Add -3 times the first row to the third row.

First, we notice that -3 times the first row gives the row vector $(-3 \ -12 \ -21)$. Then, we add this to the row vector of the third row $(3 \ 6 \ 9)$ to get the row vector $(0 \ -6 \ -12)$. Finally, this row vector replaces the third row of the matrix to give the result

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 0 & -6 & -12 \end{pmatrix}$$

Note: Notice that by performing this row sum and replacement, we have made the first column of the third row of **A** a 0. This zeroing out of the columns will play a very important role moving forward.

Using R, this can be done as

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
# extract the rows
row1 <- A[1, ]
row3 <- A[3, ]
A[3, ] <- -3 * row1 + row3
A

##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    0   -6   -12
```

Another way to do this row sum and replacement is

```
A <- matrix(c(1:9), 3, 3, byrow = FALSE)
A[3, ] <- -3 * A[1, ] + A[3, ]
A

##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    0   -6   -12
```

- 3) Multiply the second row by $\frac{1}{2}$.

The second row of \mathbf{A} is the row vector $(2 \ 5 \ 8)$. Multiplying the second row by $\frac{1}{2}$ gives the row vector $(1 \ 5/2 \ 4)$. Plugging this into the matrix gives the full matrix with second row multiplied by $\frac{1}{2}$ as

$$\begin{pmatrix} 1 & 4 & 7 \\ 1 & 5/2 & 4 \\ 3 & 6 & 9 \end{pmatrix}$$

In R, the second row of \mathbf{A} can be multiplied by $\frac{1}{2}$ as

```
A[2,] <- 1/2 * A[2,]
A
```

```
##      [,1] [,2] [,3]
## [1,]     1   4.0   7
## [2,]     1   2.5   4
## [3,]     3   6.0   9
```

2.1.6 The Augmented matrix form of a system of equations

Consider the linear system of equations

$$\begin{aligned} x_1 + 4x_2 - x_3 &= 11 \\ 4x_1 + 5x_2 + 2x_3 &= 9. \end{aligned}$$

The augmented matrix representation of this system of linear equations is given by the matrix

$$\begin{pmatrix} 1 & 4 & -1 & 11 \\ 4 & 5 & 2 & 9 \end{pmatrix},$$

where the first column of the matrix represents the variable x_1 , the second column of the matrix represents the variable x_2 , the third column of the matrix represents the variable x_3 , and the fourth column of the matrix represents the constant terms. We can express the augmented form in R using a matrix

```
augmented_matrix <- matrix(c(1, 4, 4, 5, -1, 2, 11, 9), 2, 4)
augmented_matrix
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     1     4    -1   11
## [2,]     4     5     2    9
```

and to make clear the respective variables, we can add in column names as a matrix attribute using the `colnames()` function

```
colnames(augmented_matrix) <- c("x1", "x2", "x3", "constants")
augmented_matrix
```

```
##      x1 x2 x3 constants
## [1,]  1  4 -1       11
## [2,]  4  5  2        9
```

which adds labels to each of the columns.

Now, using elementary row operations on the matrix, we can attempt to find solutions to the system of equations. First, we multiply the first row by -4 and add it to the second row of the matrix and replace the second row with this sum

```
augmented_matrix[2, ] <- -4 * augmented_matrix[1, ] + augmented_matrix[2, ]
augmented_matrix
```

```
##      x1 x2 x3 constants
## [1,]  1  4 -1       11
## [2,]  0 -11  6      -35
```

Next, scale the second row to have a leading value of 1 by dividing by -11

```
augmented_matrix[2, ] <- augmented_matrix[2, ] / (-11)
augmented_matrix
```

```
##      x1 x2           x3 constants
## [1,]  1  4 -1.0000000 11.000000
## [2,]  0  1 -0.5454545  3.181818
```

We can then multiply the second row by -4 and add it to the first row and replace the first row with this value.

```
augmented_matrix[1, ] <- augmented_matrix[1, ] - 4 * augmented_matrix[2, ]
augmented_matrix
```

```
##      x1 x2           x3 constants
## [1,]  1  0  1.1818182 -1.727273
## [2,]  0  1 -0.5454545  3.181818
```

Notice how the matrix has a “triangular” form (The lower part of the “triangle” is made of 0s and the upper part has numbers).

The triangular form tells us that There are infinitely many solutions to this system of equation. The infinite solutions are subject to the requirements that

$$x_1 = -\frac{19}{11} - \frac{13}{11}x_3$$

and

$$x_2 = \frac{35}{11} + \frac{6}{11}x_3.$$

To get this into a reasonable form, we will solve these equations as a function of x_1 . Solving the first equation for x_3 gives

$$x_3 = -\frac{19}{13} - \frac{11}{13}x_1.$$

Then, plugging this into x_3 in the second equation gives

$$\begin{aligned} x_2 &= \frac{35}{11} + \frac{6}{11} \left(-\frac{19}{13} - \frac{11}{13}x_1 \right) \\ &= \frac{341}{143} - \frac{6}{13}x_1 \end{aligned}$$

which defines a linear relationship between x_1 and x_2 . Notice that in these last two solutions, x_1 is a “free variable” and x_2 and x_3 are “determined” by x_1 .

In the plot below, the two planes (red and blue) are the geometric plots of the linear equations in the system of equations (the red plane is the top equation and the blue plane is the bottom equation). The purple line is the equation for the solution given the free variable x_3 and lies at the intersection of the two planes, much like the point in the two lines in figure [linking reference here](#) lies at the intersection of the two points.

```
# uses gg3D library
n <- 60
x1 <- x2 <- seq(-10, 10, length = n)
region <- expand.grid(x1 = x1, x2 = x2)
df <- data.frame(
  x1 = region$x1,
  x2 = region$x2,
  x3 = - 11 + (region$x1 + 4 * region$x2)
)

df2 <- data.frame(
  x1 = region$x1,
  x2 = region$x2,
  x3 = (9 - 4 * region$x1 - 5 * region$x2) / 2
```

```

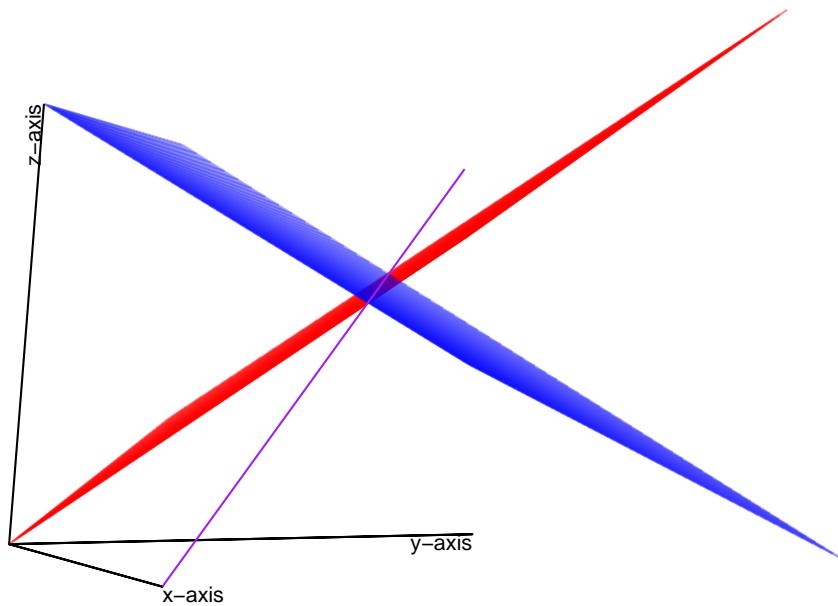
)
df_solution <- data.frame(
  x1 = x1,
  x2 = 341 / 143 - 6 / 13 * x1,
  x3 = -19/13 - 11/13 * x1
)

# theta and phi set up the "perspective/viewing angle" of the 3D plot
theta <- 63
phi <- -12
ggplot(df, aes(x1, x2, z = x3)) +
  axes_3D(theta = theta, phi = phi) +
  stat_wireframe(alpha = 0.25, color = "red", theta = theta, phi = phi) +
  stat_wireframe(data = df2, aes(x = x1, y = x2, z = x3), alpha = 0.25, color = "blue")
  stat_3D(data = df_solution, aes(x1, x2, z = x3), geom = "line", theta = theta, phi = phi)
  theme_void() +
  theme(legend.position = "none") +
  labs_3D(hjust=c(0,1,1), vjust=c(1, 1, -0.2), angle=c(0, 0, 90), theta = theta, phi = phi)
}

## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).

```



2.1.7 Existence and Uniqueness

Definition 2.6. A system of linear equations is said to be consistent if at least one solution exists. The linear system of equations is said to have a **unique** solution if only one solution exists.

Example 2.15. Is the system of linear equations consistent? If the system is consistent, does it have a unique solution?

$$\begin{aligned} 16x_1 + 2x_2 + 3x_3 &= 13 \\ 5x_1 + 11x_2 + 10x_3 &= 8 \\ 9x_1 + 7x_2 + 6x_3 &= 12 \\ 4x_1 + 14x_2 + 15x_3 &= 1 \end{aligned}$$

Solution. fill in solution here

Example 2.16. Is the system of linear equations consistent? If the system is consistent, does it have a unique solution?

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 5 \\ x_1 + 3x_2 + 2x_3 &= 2 \\ 3x_1 + 2x_2 + x_3 &= 7 \end{aligned}$$

Solution. fill in solution here

2.2 Reduce row echelon form

Reducing a matrix to row echelon form is a useful technique for working with matrices. The row echelon form can be used to solve systems of equations, as well as determine other properties of a matrix that are yet to be discussed, including rank, invertibility, column/row spaces, etc.

Definition 2.7. A matrix is said to be in **echelon** form if

- 1) all nonzero rows are above any rows of zeros (all rows consisting entirely of zeros are at the bottom)
- 2) the leading entry/coefficient of a nonzero row (called the **pivot**) is always *strictly* to the right of the leading entry/coefficient of the row above

Example 2.17. echelon matrix example in class

Definition 2.8. A matrix is in **reduced row echelon form** if it is in echelon form and

- 1) the leading entry/coefficient of each row is 1
- 2) The leading entry/coefficient of 1 is the only nonzero entry in its column.

Example 2.18. rref matrix example in class

Definition 2.9. Echelon matrices have the property of being **upper diagonal**. A matrix is said to be **upper diagonal** if all entries of the matrix at or above the diagonal are nonzero.

- Example: **lower and non-lower diagonal matrices

Definition 2.10. Two matrices are **row-equivalent** if one matrix can be transformed to the other through elementary row operations.

Theorem 2.1. *A nonzero matrix can be transformed into more than one echelon forms. However, the reduced row echelon form of a nonzero matrix is unique.*

:::{.exercise}

Using elementary row operations, calculate the reduced row echelon form of the following matrices 1) **fill in later** 2) **fill in later** 3) **fill in later**

:::

2.2.1 Pivot positions

The leading entry/coefficients of a row echelon form matrix are called pivots. The positions of the pivot positions are the same for any row echelon form of a matrix. In reduced row echelon form, these pivot positions take the value 1.

Definition 2.11. In a matrix that is in reduced echelon form, the pivot position is the first nonzero element of each row. The column in which the pivot position occurs is called a pivot column.

Example 2.19. pivot position and pivot columns

2.2.2 Finding the reduced row echelon form

Calculating the reduced row echelon form is known as Gaussian elimination, which is named after Johann Carl Friedrich Gauss. This algorithm uses elementary row operations to calculate the reduced row echelon form. The following steps perform the Gaussian elimination algorithm.

- 1) Start with the left-most nonzero column, which is a pivot column
- 2) If the top row is zero, swap rows so that the top row is nonzero so that the top row has a nonzero element in the pivot position.

- 3) Use row multiplication and addition to zero out all positions in the pivot column below the top row (pivot position).
- 4) Ignore this top row and repeat steps 1-3 until there are no more nonzero rows to apply steps 1-3 on. At the end of this step, the matrix is in row echelon form.
- 5) Starting at the right-most pivot column, use elementary row operations to zero out all positions above each pivot and to make each pivot position 1. At the end of this step, the matrix is in reduced row echelon form.

Example 2.20. in class

```
# pracma library
# rref example in class
```

2.2.3 Using reduced row echelon forms to solve systems of linear equations

When a system of linear equations is expressed as an augmented matrix, the reduced row echelon form can be used to find solutions to those systems of equations. Consider the systems of equations

$$\begin{aligned} 3x_1 + 8x_2 - 4x_3 &= 6 \\ 2x_1 - 4x_2 - 1x_3 &= 8 \\ 4x_1 + 5x_2 &= 9 \end{aligned}$$

which can be written in the augmented matrix form as

$$\begin{pmatrix} 3 & 8 & -4 & 6 \\ 2 & -4 & -1 & 8 \\ 4 & 5 & 0 & 9 \end{pmatrix}$$

In R, this is the matrix

```
# define matrix
```

Calculating the reduced row echelon form, gives

```
# calculate rref of augmented matrix
```

which gives the solution ...

Example 2.21. • calculate the RREF for the augmented matrix in the example above by hand

Example 2.22. Another example where we find a solution is

$$\begin{array}{rcl} 5x_1 & + & 4x_2 & - & 2x_3 & = 0 \\ -3x_1 & - & 2x_2 & - & 4x_3 & = 1 \end{array}$$

Do same steps

Definition 2.12. In a system of linear equations that is underdetermined (fewer equations than unknowns), the **determined/basic variables** are those variable that have a 1 in the respective columns when in reduced row echelon form (i.e., variables in a pivot position). The variables that are not in a pivot position are called **free variable**.

Example 2.23. in class

2.2.4 Existence and uniqueness from reduced row echelon form

The row echelon form is useful to determine if a system of linear equations is consistent (the system of equations has a solution). To check if a solution to a linear system of equations exists, convert the system of equations to an augmented matrix form. Then, reduce the augmented matrix to row echelon form using elementary matrix operations. As long as there is not an equation of the form

$$0 = \text{constant}$$

for some constant number not equal to 0, the system of linear equations is consistent. If the augmented matrix can be written in reduced row echelon form with no free variables, the solution to the linear system of equations is unique. These results give rise to the theorem

Theorem 2.2. *A linear system of equations is consistent (has a solution) if the furthest right column (the constant column) is not a pivot column. If the system of equations is consistent, (i.e., the furthest right column is not a pivot column), the solution is unique if there are no free variables and there are infinitely many solutions if there is at least one free variable.*

- **Example: consistent system of equations**

$$\left(\begin{array}{cccc} -7 & -9 & 7 & 8 \\ -4 & 0 & 6 & -6 \\ -10 & 3 & -8 & 5 \end{array} \right)$$

- **Example: inconsistent system of equations**

$$\begin{pmatrix} -7 & 0 & -8 & -5 \\ -4 & 3 & 8 & -2 \\ -10 & 7 & -6 & 4 \\ -9 & 6 & 5 & 1 \end{pmatrix}$$

Chapter 3

Vectors spaces

- 3 Blue 1 Brown – Linear combinations, span, and basis vectors

```
library(shiny)
library(patchwork)
library(tidyverse)
# if gg3D package not installed, install the package
library(gg3D)
library(dasc2594)
```

3.1 Vectors

3.1.1 Properties of Vectors

For any real valued scalars $a, b \in \mathcal{R}$ and any vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{R}^n$ (vectors of real numbers of length n),

1) scalar multiplication

$$\begin{aligned} a\mathbf{x} &= a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix} \end{aligned}$$

where the scalar a is multiplied by each element of the vector. For example,

$$\begin{aligned} 4 \begin{pmatrix} 4 \\ 6 \\ 7 \\ 12 \end{pmatrix} &= \begin{pmatrix} 4 * 4 \\ 4 * 6 \\ 4 * 7 \\ 4 * 12 \end{pmatrix} \\ &= \begin{pmatrix} 16 \\ 24 \\ 28 \\ 48 \end{pmatrix} \end{aligned}$$

In R, we can multiply the vector by a scalar as

```
4 * c(4, 6, 7, 12)
```

```
## [1] 16 24 28 48
```

or if the vector $\mathbf{x} = (4, 6, 7, 12)'$ we can write this as

```
x <- c(4, 6, 7, 12)
4 * x
```

```
## [1] 16 24 28 48
```

2) scalar multiplicative commutativity

$$a(b\mathbf{x}) = (ab)\mathbf{x} = b(a\mathbf{x})$$

```
4 * (6 * x)
```

```
## [1] 96 144 168 288
```

```
(4 * 6) * x
```

```
## [1] 96 144 168 288
```

3) scalar additive associativity

$$a\mathbf{x} + b\mathbf{x} = (a + b)\mathbf{x}$$

4) vector additive associativity

$$a\mathbf{x} + a\mathbf{y} = a(\mathbf{x} + \mathbf{y})$$

5) vector associativity

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$$

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$$

```
x <- c(1, 2, 3, 4)
y <- c(4, 3, 5, 1)
z <- c(5, 2, 4, 6)

x + y
```

```
## [1] 5 5 8 5
```

```
y + x
```

```
## [1] 5 5 8 5
```

```
(x + y) + z
```

```
## [1] 10 7 12 11
```

```
x + (y + z)
```

```
## [1] 10 7 12 11
```

- 6) **Identity Element of Addition:** For any vector \mathbf{x} of length n , there exists a vector $\mathbf{0}$, known as the *zero vector*, such that

$$\mathbf{x} + \mathbf{0} = \mathbf{x}$$

```
x + 0
```

```
## [1] 1 2 3 4
```

```
x + rep(0, 4)
```

```
## [1] 1 2 3 4
```

- 7) **Inverse Element of Addition:** For any vector \mathbf{x} of length n , there exists a vector $-\mathbf{x}$, known as the *additive inverse* vector, such that

$$\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$$

```
x + (-x)
```

```
## [1] 0 0 0 0
```

3.2 Vector addition

Two vectors of length n can be added elementwise

$$\begin{aligned}\mathbf{x} + \mathbf{y} &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}\end{aligned}$$

For example,

$$\begin{aligned}\begin{pmatrix} 3 \\ 1 \\ -4 \\ 3 \end{pmatrix} + \begin{pmatrix} -3 \\ 17 \\ -39 \\ 4 \end{pmatrix} &= \begin{pmatrix} 3 + (-3) \\ 1 + 17 \\ -4 + (-39) \\ 3 + 4 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 18 \\ -43 \\ 7 \end{pmatrix}\end{aligned}$$

In R, we have

```
x <- c(3, 1, -4, 3)
y <- c(-3, 17, -39, 4)
x + y
```

```
## [1] 0 18 -43 7
```

If two vectors **x** and **y** are of different lengths, then they cannot be added together. Using R, we get the following error:

```
x <- c(1, 2, 3)
y <- c(1, 2, 3, 4)
x + y
```

```
## Warning in x + y: longer object length is not a multiple of shorter object
## length
```

```
## [1] 2 4 6 5
```

The error is telling us that the vector **x** and the vector **y** do not have the same length.

Be careful when adding vectors in R. R uses “recycling” which means two vectors of different lengths can be added together if one vector is of a length that is a multiple of the other vector. For example, if **x** = (1,2)' is a vector of length 2 and **y** = (1,2,3,4) is a vector of length 4, R will add **x** + **y** by replicating the vector **x** twice (i.e., **x** + **y** = (**x**',**x**')' = (1,2,1,2)' + **y**)

```
x <- c(1, 2)
y <- c(1, 2, 3, 4)
x + y
```

```
## [1] 2 4 4 6
```

```
# replicated x = c(1, 2, 1, 2)
c(1, 2, 1, 2) + y
```

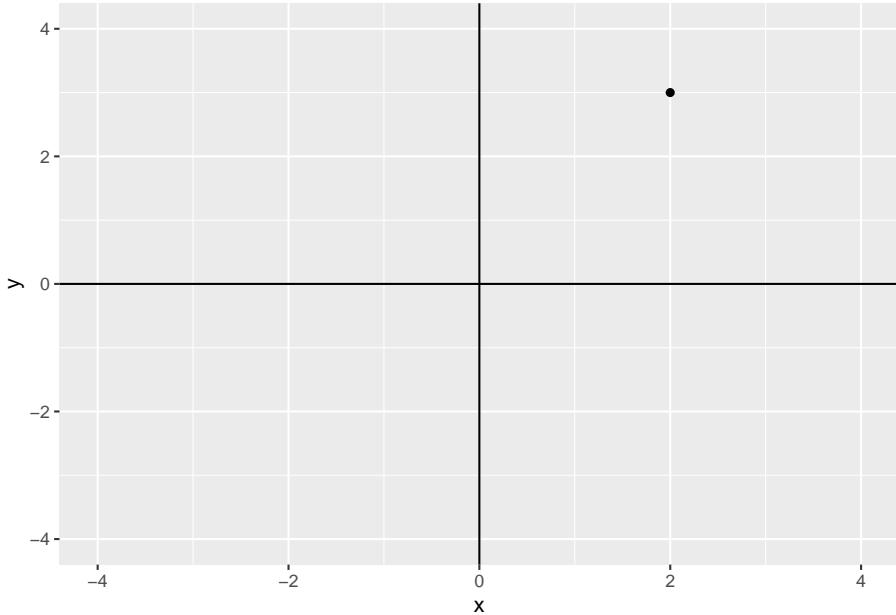
```
## [1] 2 4 4 6
```

3.2.1 The geometric interpretation of vectors in \mathcal{R}^2

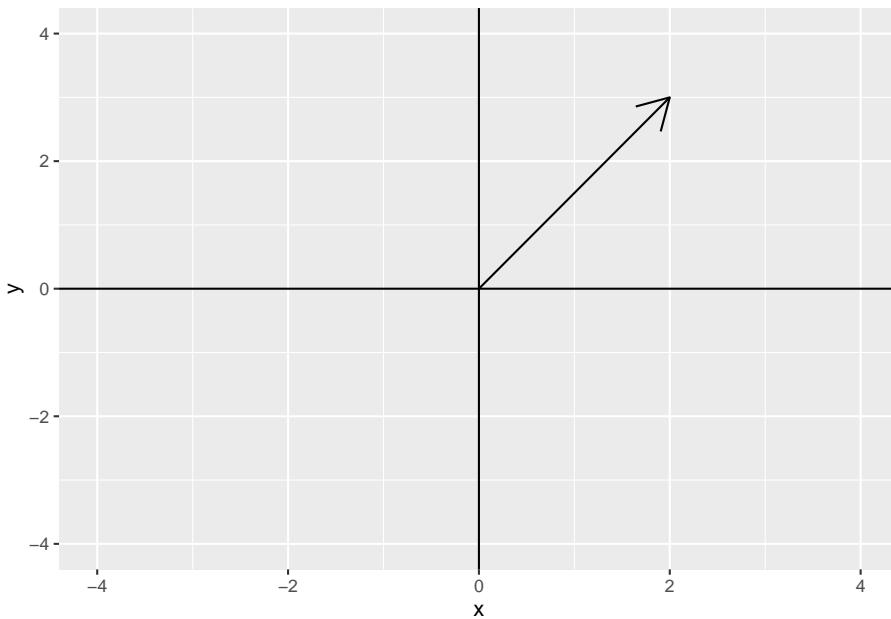
Let \mathcal{R}^2 be a real coordinate space of 2 dimensions. You are already familiar with the Cartesian plane that consists of ordered pairs (x, y) . The Cartesian plane defines the real coordinate space \mathbf{R}^2 of two dimensions. In \mathbf{R}^2 , the location of any point of interest can be defined using the x and y . For example, the plot below shows the location of the point $(2, 3)$

```
dat <- data.frame(
  x = 2,
  y = 3
)

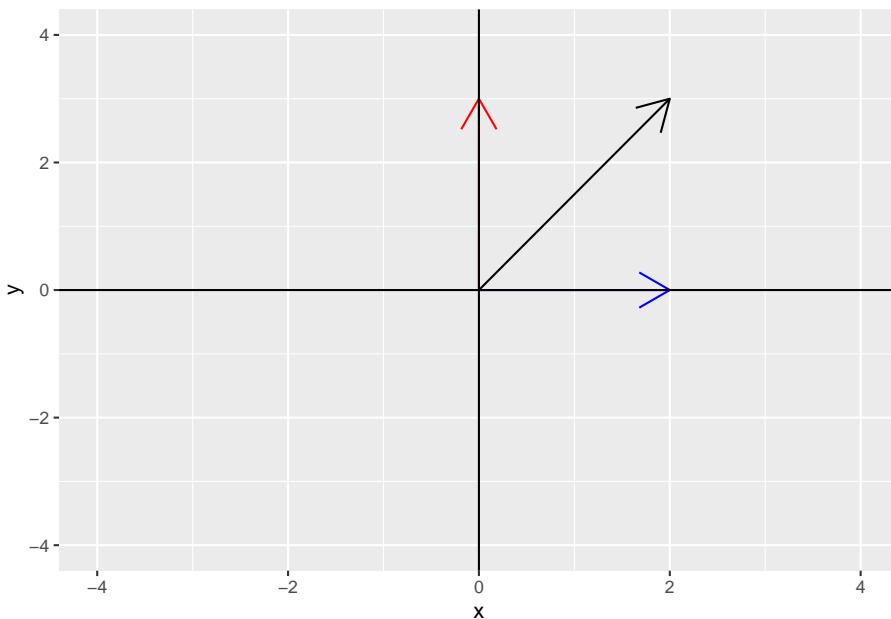
ggplot(data = dat, aes(x = x, y = y)) +
  geom_point() +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4))
```



A vector space is a generalization of this representation. In \mathcal{R}^2 , we say that the vector $\mathbf{z} = c(2, 3)$ is centered at the origin $(0, 0)$ and has length 2 in the x -axis and length 3 in the y -axis. The plot below shows this vector



We can also decompose the vector \mathbf{z} into its x and y components. The x component of \mathbf{z} is $(2, 0)$ and the y component of \mathbf{z} is $(0, 3)$. The following plot shows the x component $(2, 0)$ in blue and the y component $(0, 3)$ in red.



The below Shiny app allows you to plot the vector for any (x, y) pair of your

choosing.

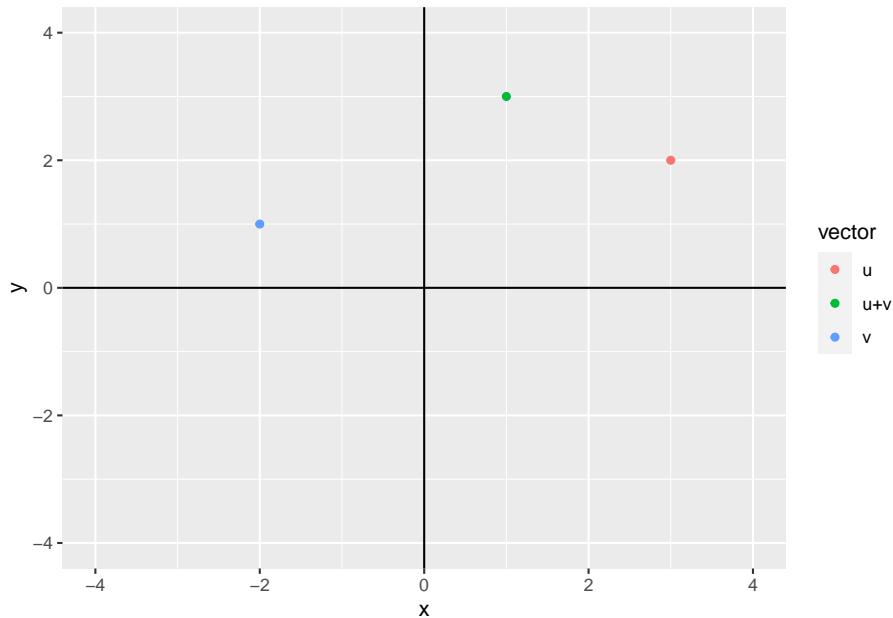
The shiny app can be downloaded and run on your own computer using

```
library(shiny)
runGitHub(rep = "multivariable-math",
          username = "jtipton25",
          subdir = "shiny-apps/chapter-03/vector-space")
```

3.2.1.1 Addition of vectors

We can represent the addition of vectors geometrically as well. Consider the two vectors $\mathbf{u} = (3, 2)$ and $\mathbf{v} = (-2, 1)$ where $\mathbf{u} + \mathbf{v} = (1, 3)$.

```
data.frame(x = c(3, -2, 1), y = c(2, 1, 3), vector = c("u", "v", "u+v")) %>%
  ggplot() +
  geom_point(aes(x = x, y = y, color = vector)) +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4))
```

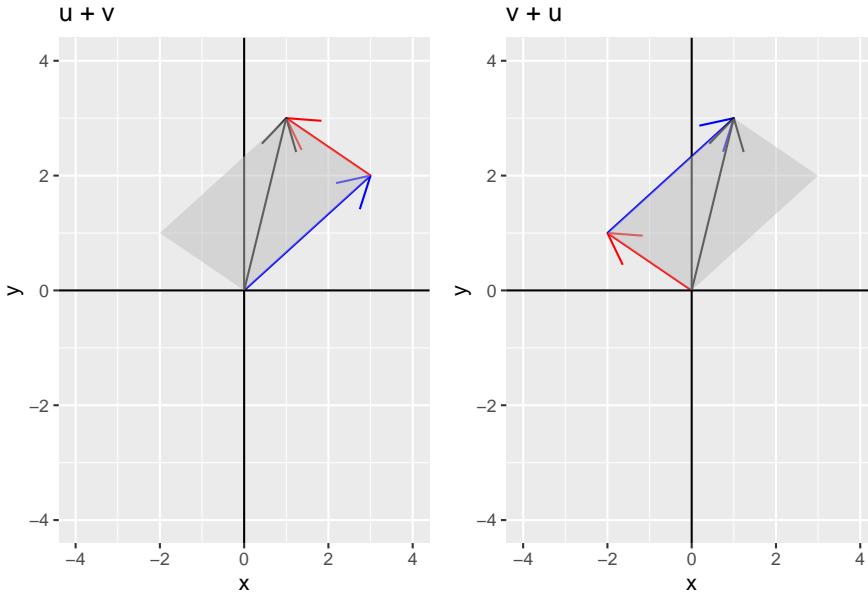


We can represent the sum using vectors by adding \mathbf{u} first then adding \mathbf{v} to \mathbf{u} or by adding \mathbf{v} first and then \mathbf{u} to get

```

df <- data.frame(x = c(0, 3, 1, -2), y = c(0, 2, 3, 1))
p1 <- ggplot() +
  geom_segment(aes(x = 0, xend = 3, y = 0, yend = 2), arrow = arrow(), color = "blue") +
  geom_segment(aes(x = 3, xend = 3 - 2, y = 2, yend = 2 + 1), arrow = arrow(), color = "red") +
  geom_segment(aes(x = 0, xend = 3 - 2, y = 0, yend = 2 + 1), arrow = arrow(), color = "black") +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_polygon(data = df, aes(x = x, y = y), fill = "grey", alpha = 0.5) +
  ggtitle("u + v")
p2 <- ggplot() +
  geom_segment(aes(x = 0, xend = -2, y = 0, yend = 1), arrow = arrow(), color = "red") +
  geom_segment(aes(x = -2, xend = -2 + 3, y = 1, yend = 1 + 2), arrow = arrow(), color = "blue") +
  geom_segment(aes(x = 0, xend = 3 - 2, y = 0, yend = 2 + 1), arrow = arrow(), color = "black") +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_polygon(data = df, aes(x = x, y = y), fill = "grey", alpha = 0.5) +
  ggtitle("v + u")
p1 + p2

```

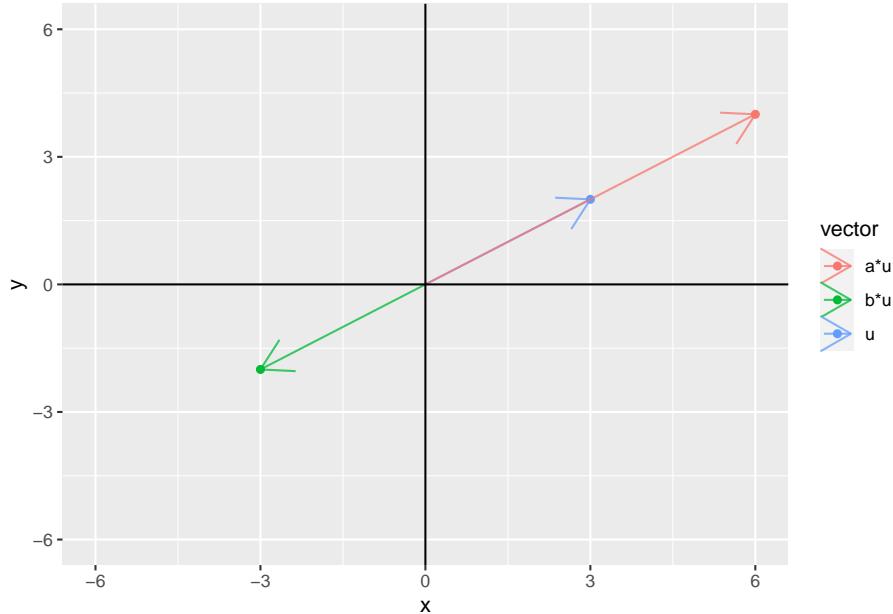


Notice that the sum of these vectors defines a parallelogram where the sum $\mathbf{u}+\mathbf{v}$ is the diagonal of the shaded parallelogram. This geometric interpretation will serve as a basis for interpreting vector equations in higher dimensions where typical visualization methods fail.

3.2.2 Scalar multiplication of vectors

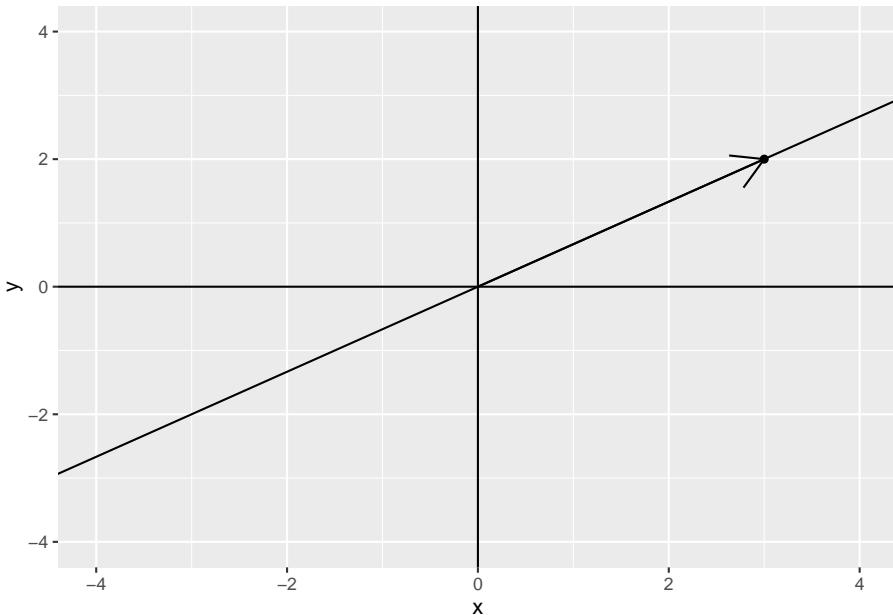
We can represent the multiplication of a vector by a scalar geometrically as well. Consider the vector $\mathbf{u} = (3, 2)$ and the scalars $a = 2$ and $b = -1$. Then, we can plot \mathbf{u} , $a\mathbf{u}$, and $b\mathbf{u}$.

```
data.frame(x = c(3, 2 * 3, -1 * 3), y = c(2, 2 * 2, -1 * 2), vector = c("u", "a*u", "b*u"))
ggplot() +
  geom_point(aes(x = x, y = y, color = vector)) +
  geom_segment(aes(x = 0, xend = x, y = 0, yend = y, color = vector), arrow = arrow()) +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-6, 6), ylim = c(-6, 6))
```



In fact, if a is allowed to take on any values, then the set of all possible values of $a\mathbf{u}$ for all values of a defines an infinite line

```
ggplot() +
  geom_abline(slope = 2/3, intercept = 0) +
  geom_point(aes(x = 3, y = 2)) +
  geom_segment(aes(x = 0, xend = 3, y = 0, yend = 2), arrow = arrow(), color = "black") +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4))
```



3.2.3 The geometric interpretation of vectors in \mathcal{R}^3

Let the vector $\mathbf{u} = c(-2, 3, 5)$. Then, the figure below shows the vector in 3 dimensions.

Draw picture by hand

3.2.4 The geometric interpretation of vectors in \mathcal{R}^n

As the number of dimensions increases, the same interpretation can be used, but the ability to visualize higher dimensions becomes more difficult.

3.2.5 Linear Combinations of Vectors

We say that for any two scalars a and b and any two vectors \mathbf{x} and \mathbf{y} of length n , the sum

$$a\mathbf{x} + b\mathbf{y} = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ \vdots \\ ax_n + by_n \end{pmatrix}$$

is called a linear combination. The idea of a linear combination can be extended to K different scalars $\{a_1, \dots, a_K\}$ and K different vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_K\}$ each of length n as

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_K\mathbf{x}_K = \sum_{k=1}^K a_k\mathbf{x}_k = \begin{pmatrix} \sum_{k=1}^K a_k x_{k1} \\ \sum_{k=1}^K a_k x_{k2} \\ \vdots \\ \sum_{k=1}^K a_k x_{kn} \end{pmatrix}$$

The scalars a_k are called **coefficients** (sometimes also called **weights**).

- **Example:** Consider the linear combination $a\mathbf{u} + b\mathbf{v}$ where

$$\mathbf{u} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Are there values of a and b such $a\mathbf{u} + b\mathbf{v} = \begin{pmatrix} 9 \\ -4 \end{pmatrix}$? To answer this question, we can write the linear combination as

$$a \begin{pmatrix} 3 \\ 6 \end{pmatrix} + b \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ -4 \end{pmatrix}$$

which can be written using the property of scalar multiplication as

$$\begin{pmatrix} 3a \\ 6a \end{pmatrix} + \begin{pmatrix} -2b \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ -4 \end{pmatrix}$$

and using properties of vector addition can be written as

$$\begin{pmatrix} 3a - 2b \\ 6a + b \end{pmatrix} = \begin{pmatrix} 9 \\ -4 \end{pmatrix}$$

Recognizing this as a system of linear equations

$$\begin{aligned} 3a - 2b &= 9 \\ 6a + b &= -4, \end{aligned}$$

the system of equations can be written in an augmented matrix form as

$$\begin{pmatrix} 3 & -2 & 9 \\ 6 & 1 & -4 \end{pmatrix}$$

Reducing the augmented matrix to reduced row echelon form gives

```
rref(matrix(c(3, 6, -2, 1, 9, -4), 2, 3))
```

```
##      [,1] [,2]      [,3]
## [1,]    1    0  0.06666667
## [2,]    0    1 -4.40000000
```

which has solutions $a = 0.0667$ and $b = -4.4$.

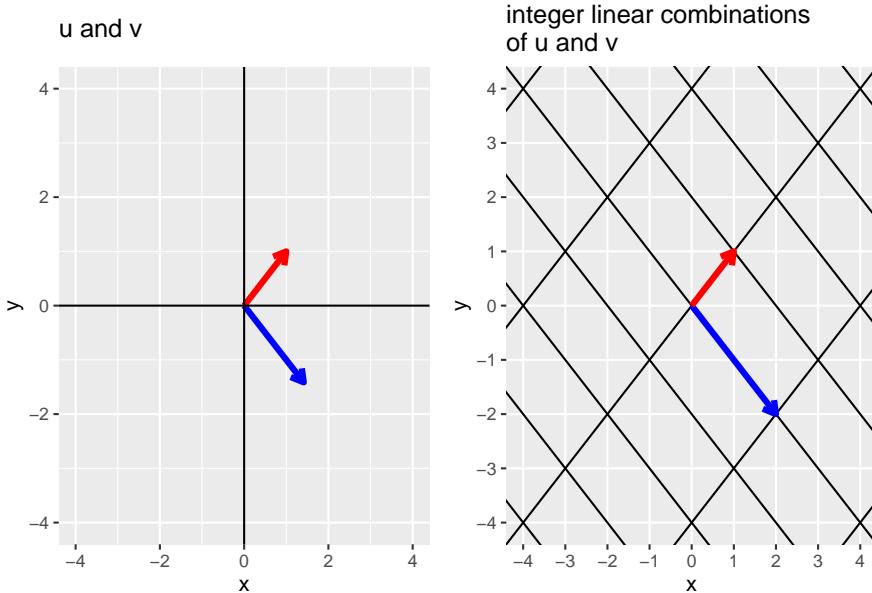
- **Result:** Any vector equation $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_K\mathbf{x}_K = \mathbf{c}$ for a given constant vector \mathbf{b} has the same solution set as the augmented matrix

$$(\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_K \quad \mathbf{b})$$

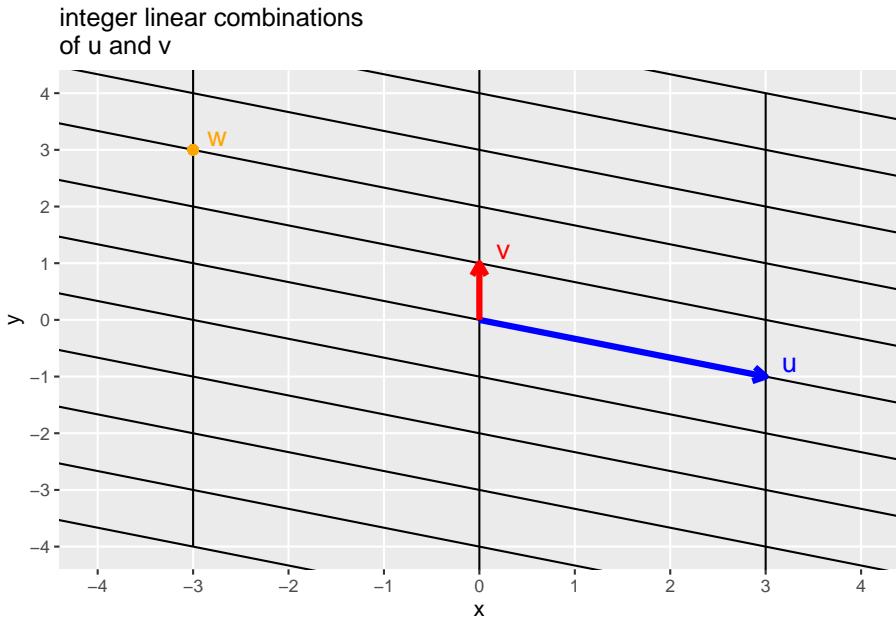
Equivalently, the set of vectors $\{\mathbf{x}_k\}_{k=1}^K$ can only be combined with linear coefficients $\{a_k\}_{k=1}^K$ to equal the vector \mathbf{b} if the linear system of equations is consistent.

3.2.6 The geometric interpretation of linear combinations of vectors

Consider the vectors $\mathbf{u} = \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ shown in the figure below on the left.



- **Exercise:** Given $\mathbf{u} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, estimate the linear combination of \mathbf{u} and \mathbf{v} that gives the point \mathbf{w} in the figure below.



3.3 Span

Definition 3.1 (Span). Let $\mathbf{a}_1, \dots, \mathbf{a}_K$ be vectors in \mathcal{R}^n . We say the vector \mathbf{w} is in the span of $\mathbf{a}_1, \dots, \mathbf{a}_K$ ($\mathbf{w} \in \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_K\}$) if there exists coefficients x_1, \dots, x_K such that $\mathbf{w} = \sum_{k=1}^K x_k \mathbf{a}_k$.

Example 3.1. While not a vector notation, you already understand the span from polynomial functions. For example, assume you have the functions 1, x , and x^2 . Then, the functions $-4 + 3x^2$ ($a_1 = -4, a_2 = 0, a_3 = 3$) and $-3 + 4x - 2x^2$ ($a_1 = -3, a_2 = 4, a_3 = -2$) are in the span of functions $\{1, x, x^2\}$, but the functions $x^3, x^4 - 2x^2$, etc., are not in the span of $\{1, x, x^2\}$ because you cannot write these as a linear combination of $a_1 1 + a_2 x + a_3 x^2$.

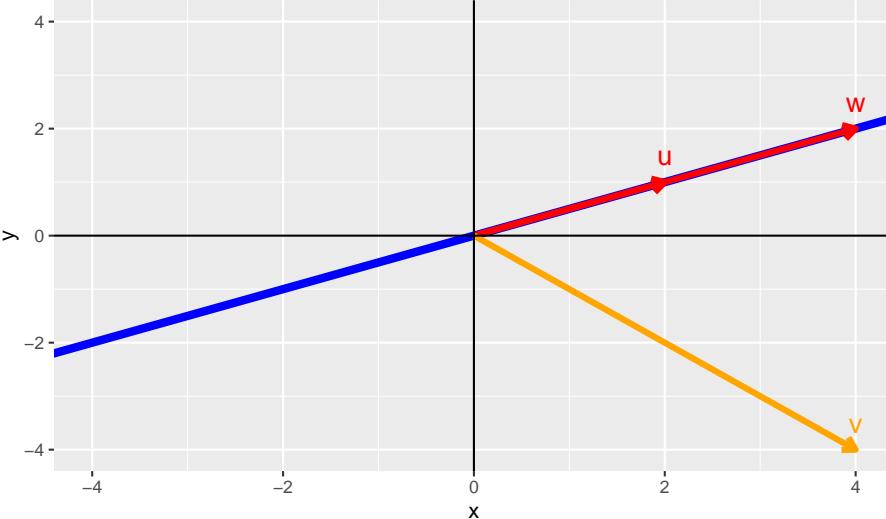
3.3.1 Geometric example of the span

- **Example:** Consider the vector $\mathbf{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Then, the vector $\mathbf{w} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ is in the span $\{\mathbf{u}\}$ because $\mathbf{w} = 2\mathbf{u}$ but the vector $\mathbf{v} = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$ is not in

the $\text{span}\{\mathbf{u}\}$ because there is no coefficient a such that $\mathbf{w} = a\mathbf{u}$. In this example, the vector \mathbf{u} is a 2-dimensional vector (lives in \mathcal{R}^2 —a plane) but the $\text{span}\{\mathbf{u}\}$ lives in 1-dimension (a line).

```
ggplot() +
  geom_abline(slope = 1/2, intercept = 0, color = "blue", size = 2) +
  geom_segment(aes(x = 0, xend = 2, y = 0, yend = 1), arrow = arrow(length = unit(0.1, "inches")))
  geom_segment(aes(x = 0, xend = 4, y = 0, yend = 2), arrow = arrow(length = unit(0.1, "inches")))
  geom_segment(aes(x = 0, xend = 4, y = 0, yend = -4), arrow = arrow(length = unit(0.1, "inches")))
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_text(data = data.frame(x = c(2, 4, 4), y = c(1, 2, -4), text = c("u", "w", "v")),
            aes(x = x, y = y + 0.5, label = text), size = 5, inherit.aes = FALSE,
            color = c("red", "red", "orange")) +
  ggttitle("span{u} is the blue line \nw is in span{u}\nv is not in span{u}")
```

span{u} is the blue line
 w is in span{u}
 v is not in span{u}



From the example above, we can answer the question “Is the point (a, b) on the line defined by the vector \mathbf{u} ?” by asking whether the point (a, b) is in the $\text{span}\{\mathbf{u}\}$. While this is trivial for such a simple problem, the use of the span will make things easier in higher dimensions.

- **Example: do in class** 2 3-d vectors that are not scalar multiples of each other define a plane. Does a point lie within the plane? Use the span to answer this question.

The shiny app below demonstrates how the concept of span can be understood in 2 dimensions. The app can be downloaded and run

```
library(shiny)
runGitHub(rep = "multivariable-math",
          username = "jtipton25",
          subdir = "shiny-apps/chapter-03/span")
```

Chapter 4

Matrix equations

```
library(tidyverse)
library(dasc2594)
library(gg3D)
```

Here we introduce the concept of the linear equation $\mathbf{Ax} = \mathbf{b}$. This equation is the most fundamental equation in all of statistics and data science. Given a matrix \mathbf{A} and a vector of constants \mathbf{b} , the goal is to solve for the value (or values) of \mathbf{x} that are a solution to this equation. The equation $\mathbf{Ax} = \mathbf{b}$ is a matrix representation of the system of linear equations

$$\begin{aligned}\mathbf{Ax} &= \mathbf{b} \\ (\mathbf{a}_1 &\quad \dots \quad \mathbf{a}_K) \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix} = \mathbf{b} \\ x_1\mathbf{a}_1 + \dots + x_K\mathbf{a}_K &= \mathbf{b}\end{aligned}$$

as long as the matrix \mathbf{A} has n rows and K columns and the vectors \mathbf{a}_k are n -dimensional.

- Example: in class
- Example: in class

4.1 Solutions of matrix equations

Because the matrix equation $\mathbf{Ax} = \mathbf{b}$ is equivalent to a linear system of equations $x_1\mathbf{a}_1 + \dots + x_K\mathbf{a}_K = \mathbf{b}$, we can solve the matrix equation $\mathbf{Ax} = \mathbf{b}$ by

writing the equation in an augmented matrix form

$$(\mathbf{a}_1 \ \dots \ \mathbf{a}_K \ \mathbf{b})$$

and then reducing the matrix to reduced row echelon form. This gives rise to the theorem

Theorem 4.1. *The matrix equation $\mathbf{Ax} = \mathbf{b}$, the vector equation $x_1\mathbf{a}_1 + \dots + x_K\mathbf{a}_K = \mathbf{b}$, and the augmented matrix $(\mathbf{a}_1 \ \dots \ \mathbf{a}_K \ \mathbf{b})$ all have the same solution set.*

4.2 Existence of solutions

A solution to the matrix equation $\mathbf{Ax} = \mathbf{b}$ exists if and only if \mathbf{b} is a linear combination of the columns of \mathbf{A} . In other words, $\mathbf{Ax} = \mathbf{b}$ has a solution if and only if \mathbf{b} is in the span $\{\mathbf{a}_1, \dots, \mathbf{a}_K\}$.

- **Example: in class** Let $\mathbf{A} = \dots$ and $\mathbf{b} = \dots$. Is the matrix equation $\mathbf{Ax} = \mathbf{b}$ consistent?

Theorem 4.2. *For the $n \times K$ matrix \mathbf{A} , the following statements are equivalent:*

- For each $\mathbf{b} \in \mathbb{R}^n$, the equation $\mathbf{Ax} = \mathbf{b}$ has at least one solution*
- Each $\mathbf{b} \in \mathbb{R}^n$ is a linear combination of the columns of \mathbf{A}*
- The columns of \mathbf{A} span \mathbb{R}^n*
- \mathbf{A} has n pivot columns. (\mathbf{A} has a pivot in every row)*

4.3 Matrix-vector multiplication

To calculate \mathbf{Ax} , we need to define matrix multiplication. The equivalence between the linear systems of equations $x_1\mathbf{a}_1 + \dots + x_K\mathbf{a}_K = \mathbf{b}$ and the matrix equation \mathbf{Ax} gives a hint in how to do this. First, recall the definition of \mathbf{A} and \mathbf{x}

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nK} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

The matrix product \mathbf{Ax} is the linear system of equations

$$\begin{aligned}\mathbf{Ax} &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \dots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{nK} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1K}x_K \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2K}x_K \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nK}x_K \end{pmatrix}\end{aligned}$$

Notice that the first row of the last matrix above has the sum first row of the matrix \mathbf{A} multiplied by the corresponding elements in \mathbf{x} (i.e., first element a_{11} of the first row of \mathbf{A} times the first element x_1 of \mathbf{x} plus the second, third, fourth, etc.). Likewise, this pattern holds for the second row, and all the other rows. This gives an algorithm for evaluating the product \mathbf{Ax} .

Definition 4.1. The product \mathbf{Ax} of a $n \times K$ matrix \mathbf{A} with a K -vector \mathbf{x} is a n -vector where the i th element of \mathbf{Ax} is the sum of the i th row of \mathbf{A} times the corresponding elements of the vector \mathbf{x}

- Example: in class
- Example: in class
- Example: in R using loops
- Example: in R using %*%

4.4 Properties of matrix-vector multiplication

If \mathbf{A} is a $n \times K$ matrix, \mathbf{u} and \mathbf{v} are vectors in \mathcal{R}^K and c is a scalar, then

- a) $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}$
- b) $\mathbf{A}(c\mathbf{u}) = (c\mathbf{A})\mathbf{u}$

- Proof in class

4.5 Solutions of linear systems

4.5.1 Homogeneous linear systems of equations

Definition 4.2. The matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

is called a homogeneous system of equations. The vector $\mathbf{0}$ is a vector of length N composed of all zeros. The **trivial** solution of the homogeneous equation is when $\mathbf{x} = \mathbf{0}$ and is not a very useful solution. Typically one is interested in **nontrivial** solutions where $\mathbf{x} \neq \mathbf{0}$.

The homogeneous linear system of equations can be written in augmented matrix form

$$(\mathbf{a}_1 \quad \dots \quad \mathbf{a}_K \quad \mathbf{0})$$

which implies that a non-trivial solution only exists if there is a free variable. Another way of saying this is that at least one column \mathbf{A} must not be a pivot column (Note: the last column of the augmented matrix will not be a pivot column because it will be a column of zeros). If every column of \mathbf{A} were a pivot column, the reduced row echelon form of the augmented matrix would be

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

which implies the only solution is the trivial solution $\mathbf{0}$.

- **Example: in class**

$$\begin{aligned} 3x_1 - 2x_2 + 4x_3 &= 0 \\ -2x_1 + 4x_2 - 2x_3 &= 0 \\ 5x_1 - 6x_2 + 6x_3 &= 0 \end{aligned}$$

* **Example: in class**

Consider the equation

$$2x_1 + 4x_2 - x_3 = 0.$$

we can write this as

$$x_1 = -2x_2 + \frac{1}{2}x_3$$

where x_2 and x_3 are free variables. Writing this as a solution \mathbf{x} gives

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2x_2 + \frac{1}{2}x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$$

which is a linear combination of the vectors $\mathbf{u} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$. This

implies that we can write the solution $\mathbf{x} = c\mathbf{u} + d\mathbf{v}$ for scalars c and d . Therefore, the solution set \mathbf{x} is contained in the $\text{span}\{\mathbf{u}, \mathbf{v}\}$. Because the vectors \mathbf{u} and \mathbf{v} are linearly independent (they don't point in the same direction), the set of all linear combinations of $c\mathbf{u} + d\mathbf{v}$ defines a plane.

Definition 4.3. A solution set of the form $\mathbf{x} = c\mathbf{u} + d\mathbf{v}$ is called a **parametric vector solution**.

4.6 Solutions to nonhomogeneous systems

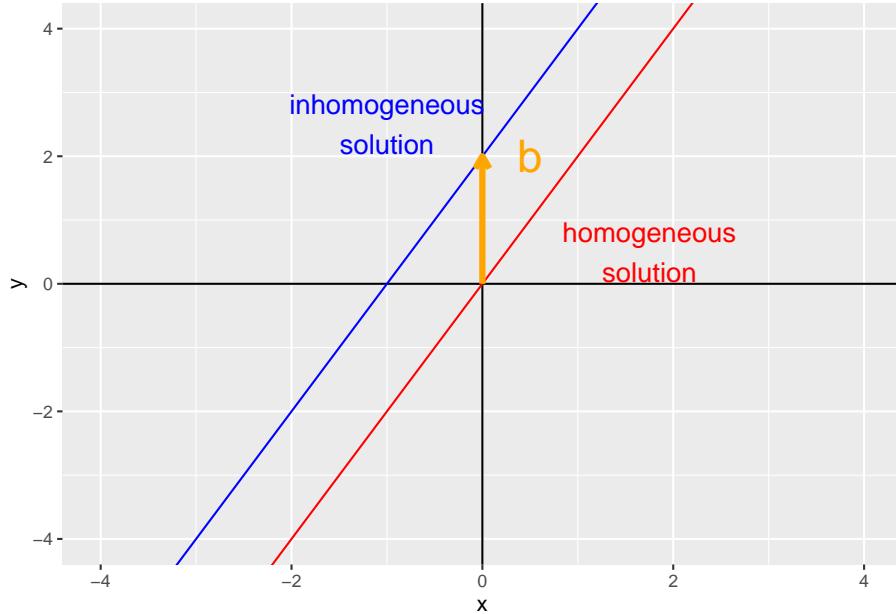
Recall the simple linear equation

$$y = mx + b$$

where m is the slope and b is the y -intercept. Setting $b = 0$ gives a simple homogenous linear equation where the y -intercept goes through the origin $(0, 0)$. When b is nonzero, the line keeps the same slope but is shifted upward/downward by b .

```
ggplot(data = data.frame(x = 0, y = 0), aes(x, y)) +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  geom_abline(slope = 2, intercept = 0, color = "red") +
  geom_abline(slope = 2, intercept = 2, color = "blue") +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_text(
    data = data.frame(x = c(0, 0), y = c(0, 2), text = c("homogeneous\\nsolution", "inhomogeneous\\nsolution")),
    aes(x = x + c(1.75, -1), y = y + 0.5, label = text), size = 5, inherit.aes = FALSE,
    color = c("red", "blue")) +
  geom_segment(
    aes(x = 0, xend = 0, y = 0, yend = 2),
    arrow = arrow(length = unit(0.1, "inches")),
    size = 1.5, color = "orange") +
  geom_text(
    data = data.frame(x = 0, y = 2, text = "b"),
    aes(x = x + 0.5, y = y, label = text)),
```

```
size = 8, inherit.aes = FALSE,
color = "orange")
```



This shift in location (but not in slope) is called a **translation**

- **Example:** Show this shift for a system of linear equations where the solution set defines a plane. From example above,

$$2x_1 + 4x_2 - x_3 = 0.$$

has the parametric solution $\mathbf{x} = c\mathbf{u} + d\mathbf{v}$ with

$$\mathbf{u} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$$

Now, if we change the system of linear equations so that we have the inhomogeneous equation

$$2x_1 + 4x_2 - x_3 = 20.$$

we get the homogeneous solution set $x_1 = -2x_2 + \frac{1}{2}x_3 + 10$ which can be written in parametric form as $\mathbf{x} = c\mathbf{u} + d\mathbf{v} + \mathbf{p}$ with

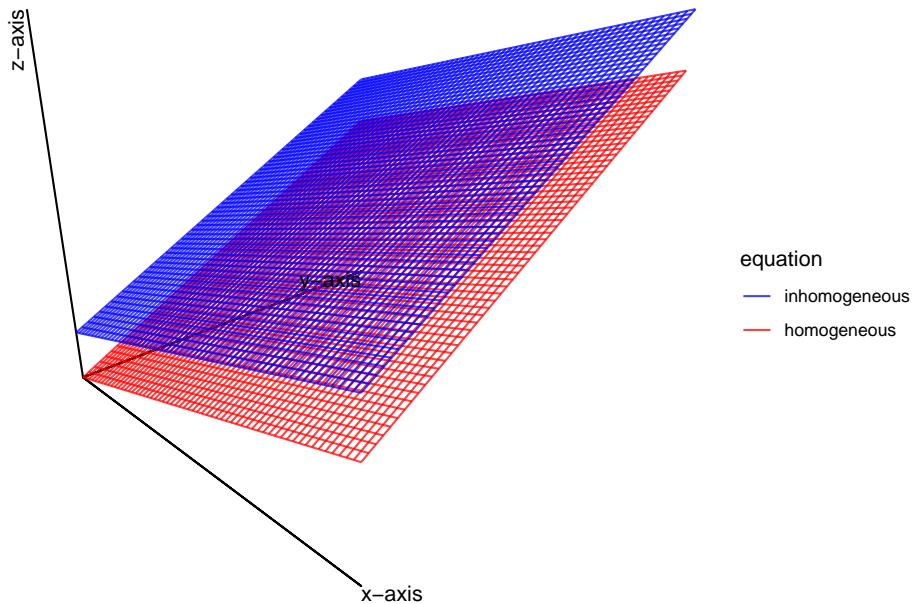
$$\mathbf{u} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} 10 \\ 0 \\ 0 \end{pmatrix}$$

For plotting, we will solve these equations for x_3 , letting x_1 and x_2 be free variables (this is just for the requirements of the plotting function). Thus, the homogeneous equation has the solution $x_3 = 2x_1 + 4x_2$ and the inhomogenous equation has the solution $x_3 = 2x_1 + 4x_2 - 20$.

```
# uses gg3D library
n <- 60
x1 <- x2 <- seq(-10, 10, length = n)
region <- expand.grid(x1 = x1, x2 = x2)
df <- data.frame(
  x1 = region$x1,
  x2 = region$x2,
  x3 = c(
    2 * region$x1 + 4 * region$x2,
    2 * region$x1 + 4 * region$x2 - 20),
  equation = rep(c("inhomogeneous", "homogeneous"), each = n^2))

# theta and phi set up the "perspective/viewing angle" of the 3D plot
theta <- 45
phi <- 20
ggplot(df, aes(x = x1, y = x2, z = x3, color = equation)) +
  axes_3D(theta = theta, phi = phi) +
  stat_wireframe(
    alpha = 0.75,
    theta = theta, phi = phi) +
  scale_color_manual(values = c("inhomogeneous" = "blue", "homogeneous" = "red")) +
  theme_void() +
  labs_3D(hjust=c(0,1,1), vjust=c(1, 1, -0.2),
         angle=c(0, 0, 90), theta = theta, phi = phi)

## Warning: Removed 4 row(s) containing missing values (geom_path).
```



- **Example: in class** Let's revisit the example from before

$$\begin{aligned} 3x_1 - 2x_2 + 4x_3 &= 0 \\ -2x_1 + 4x_2 - 2x_3 &= 0 \\ 5x_1 - 6x_2 + 6x_3 &= 0 \end{aligned}$$

but change this so that $\mathbf{b} = \begin{pmatrix} 2 \\ -6 \\ 8 \end{pmatrix}$

- Write this as a parametric solution with a mean shift

```
A <- matrix(c(3, -2, 5, -2, 4, -6, 4, -2, 6, 2, -6, 8), 3, 4)
rref(A)

##      [,1] [,2] [,3] [,4]
## [1,]    1    0  1.50 -0.50
## [2,]    0    1  0.25 -1.75
## [3,]    0    0  0.00  0.00
```

$$\begin{aligned} x_1 &= \frac{3}{2}x_2 - \frac{1}{2} \\ x_2 &= \frac{1}{4}x_3 - \frac{7}{4} \end{aligned}$$

which was the same solution set as the homogenous solution plus the additional vector $\begin{pmatrix} -\frac{1}{2} \\ -\frac{7}{4} \\ 0 \end{pmatrix}$. Thus, the inhomogenous solution is now $\mathbf{x} = c\mathbf{u} + d\mathbf{v} + \mathbf{p}$ where

$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ \frac{3}{2} \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ \frac{1}{4} \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{7}{4} \\ 0 \end{pmatrix}$$

4.7 Finding solutions

The following algorithm describes how to solve a linear system of equations.

- 1) Put the system of equations in an augmented matrix form
- 2) Reduce the augmented matrix to reduced row echelon form
- 3) Express each determined variable as a function of the free variables.
- 4) Write the solution in a general form where the determined variables are a function of the independent variables
- 5) Decompose the solution \mathbf{x} into a linear combination of free variables as parameters

Chapter 5

Linear independence

Recall the homogeneous equation $\mathbf{Ax} = \mathbf{0}$ can be written as a linear combination of coefficients x_1, \dots, x_K and vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ where

$$\sum_{k=1}^K x_k \mathbf{a}_k = \mathbf{0}$$

Definition 5.1. The set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ are called **linearly independent** if the only solution to the vector equation $\sum_{k=1}^K x_k \mathbf{a}_k = \mathbf{0}$ is the trivial solution. The set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ are called **linearly dependent** if there are coefficients x_1, \dots, x_K that are not all zero.

Example 5.1. In class

What does it mean for a set of vectors to be **linearly dependent**? This means that there is at least one vector \mathbf{a}_k that can be written as a sum of the other vectors with coefficients x_k :

$$\mathbf{a}_k = \sum_{j \neq k} x_j \mathbf{a}_j$$

Note: **linear dependence** does not imply that all vectors \mathbf{a}_k can be written as a linear combination of other vectors, just that there is at least one such vector in the set.

Example 5.2. Example: in class – determine if the vectors are linearly independent and solve the dependence relation

Theorem 5.1. *The matrix \mathbf{A} has linearly independent columns if and only if the matrix equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.*

Example 5.3. Example: in class A set of a single vector

Example 5.4. Example: in class A set of two vectors

- linearly independent if:
- linearly dependent if one vector is a scalar multiple of the other:

Theorem 5.2. *If an $n \times K$ matrix \mathbf{A} has $K > n$, then the columns of \mathbf{A} are linearly dependent. In other words, if a set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ contains more vectors than entries within vectors, the set of vectors is linearly dependent.*

Proof. If $K > n$, there are more variables (K) than equations (n). Therefore, there is at least one free variable and this implies that the homogeneous equation $\mathbf{Ax} = \mathbf{0}$ has a non-trivial solution (4.2) \square

Theorem 5.3. *If a set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ contains the $\mathbf{0}$ vector, then the the set of vectors is linearly dependent.*

Proof. in class \square

Example 5.5. In class: Determine whether the following sets of vectors are linearly dependent

Chapter 6

Linear Transformations

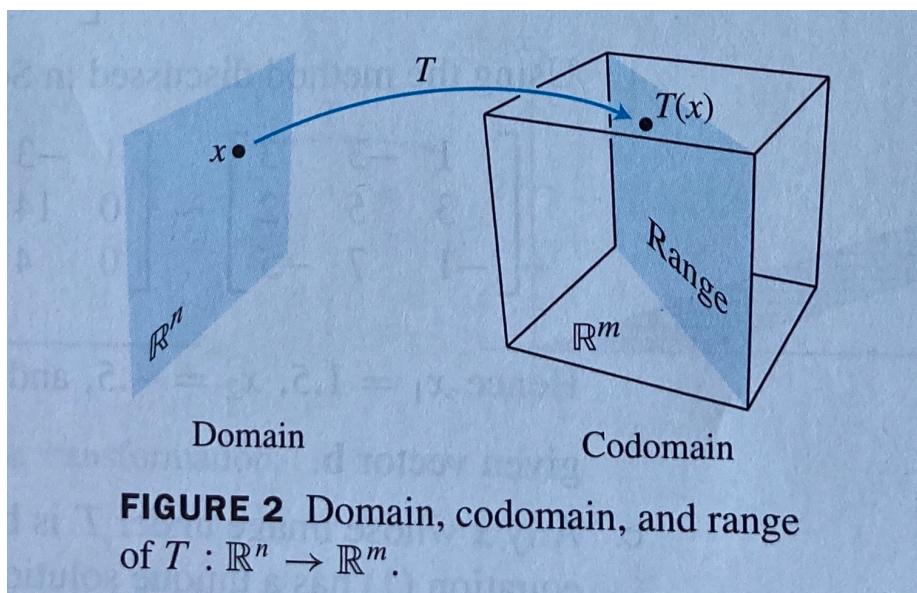
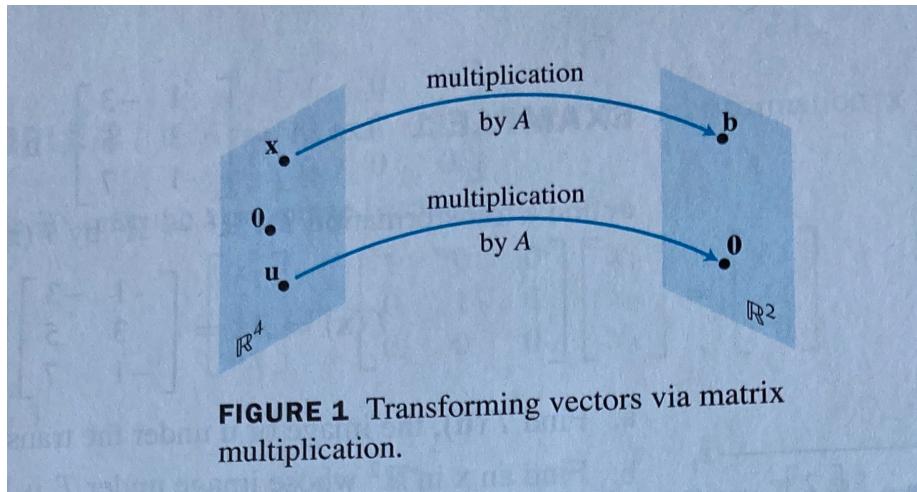
- 3 Blue 1 Brown – Linear transformations
- 3 Blue 1 Brown – 3D transformations

```
library(tidyverse)
library(dasc2594)
library(gifski)
```

It is often useful to think of \mathbf{Ax} as a linear transformation defined by the matrix \mathbf{A} applied to the vector \mathbf{x} .

A linear transformation is mathematically defined as a function/mapping $T(\cdot)$ (T for transformation) from a **domain** in \mathcal{R}^n (function input) to a **codomain** in \mathcal{R}^m (function output). In shorthand, this is written as $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ which is read a “ T maps inputs from the domain \mathcal{R}^n to the codomain \mathcal{R}^m .” For each $\mathbf{x} \in \mathcal{R}^n$ (in the domain), $T(\mathbf{x}) \in \mathcal{R}^m$ is known as the image of \mathbf{x} . The set of all $T(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{R}^n$ is known as the range of $T(\mathbf{x})$. Note that it is possible that the range of $T(\mathbf{x})$ is not required to be the entire space \mathcal{R}^m (i.e., the range of the transformation T might be a subset of \mathcal{R}^m)

Draw figure



In the case of matrix transformations (linear transformations), the function $T(\mathbf{x}) = \mathbf{Ax}$ where \mathbf{A} is a $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$ is a n -vector.

- **Question:** What kind of object is \mathbf{Ax} ?
 - scalar
 - vector
 - matrix
 - array
- **Question** What are the dimensions of \mathbf{Ax} ?

Using the matrix transformation notation, the domain of the transformation T is \mathcal{R}^n , the codomain of T \mathcal{R}^m . The range of the transformation T is the set of all linear combinations of the columns of \mathbf{A} (the $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$) because the transformation $T(\mathbf{x}) = \mathbf{Ax}$ is a linear combination $\sum_{i=1}^n x_i \mathbf{a}_i$ of the columns $\{\mathbf{a}_i\}_{i=1}^n$ of \mathbf{A} with coefficients x_1, \dots, x_n

Example 6.1.

$$\mathbf{A} = \begin{pmatrix} 2 & 4 \\ -3 & 1 \\ -1 & 6 \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} -2 \\ -11 \\ -15 \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix}$$

```
A <- matrix(c(2, -3, -1, 4, 1, 6), 3, 2)
u <- c(1, 3)
b <- 3 * A[, 1] - 2 * A[, 2]
c <- c(2, -2, -1)
```

- a) Find the image of \mathbf{u} using the matrix transformation T (e.g., calculate $T(\mathbf{u})$).

```
# a
A %*% u
```

```
##      [,1]
## [1,]    14
## [2,]     0
## [3,]    17
```

The image of \mathbf{u} under T is $T(\mathbf{u}) = \mathbf{Au} = \begin{pmatrix} 14 \\ 0 \\ 17 \end{pmatrix}$.

- b) Find a coefficient vector $\mathbf{x} \in \mathcal{R}^2$ such that $T(\mathbf{x}) = \mathbf{b}$.

```
#b
rref(cbind(A, b))
```

```
##      b
## [1,] 1 0 3
## [2,] 0 1 -2
## [3,] 0 0 0
```

A coefficient vector \mathbf{x} such that $T(\mathbf{x}) = \mathbf{b}$ is the solution to the matrix equation $\mathbf{Ax} = \mathbf{b}$ which can be found from the reduced row echelon form of the augmented matrix above giving $\mathbf{x} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$

- c) Is there more than one \mathbf{x} whose image under T is \mathbf{b} ? In other words, is the solution $\mathbf{Ax} = \mathbf{b}$ unique?

Use the reduced row echelon form of the matrix equation $\mathbf{Ax} = \mathbf{b}$ above which gives a unique solution as every (non augmented) column is a pivot. Thus, there is only one solution.

- d) Determine if \mathbf{c} is in the range of T . In other words, does the solution $\mathbf{Ax} = \mathbf{c}$ exist?

```
# d
rref(cbind(A, c)) # no because this is an inconsistent system of equations
```

```
##          c
## [1,] 1 0 0
## [2,] 0 1 0
## [3,] 0 0 1
```

The solution to the matrix equation $\mathbf{Ax} = \mathbf{c}$ which can be found from the reduced row echelon form of the augmented matrix above results in no solution because the last column (the augmented column) is a pivot column. Thus, the system of equations is inconsistent and \mathbf{c} cannot be written as a linear combination of the columns of \mathbf{A} which means that \mathbf{c} is not in the range of T .

6.1 Linear Transformations

Definition 6.1. A transformation $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is linear if

- 1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all \mathbf{u} and \mathbf{v} in the domain of T
- 2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all vectors \mathbf{u} in the domain of T

Note: Because a linear transformation is equivalent to a matrix transformation, the definition above is equivalent to the following matrix-vector multiplication properties

If \mathbf{A} is a $m \times n$ matrix, \mathbf{u} and \mathbf{v} are vectors in \mathcal{R}^m and c is a scalar, then

- 1) $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}$
- 2) $\mathbf{A}(c\mathbf{u}) = (c\mathbf{A})\mathbf{u}$

As a consequence of the previous definition, the following properties hold for scalars c and d and vectors \mathbf{u} and $\mathbf{v} \in \mathcal{R}^m$

- 3) $T(\mathbf{0}) = \mathbf{0}$
- 4) $T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$

- **Show why in class**

These properties give rise to the following statement for scalars c_1, \dots, c_m and vectors $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathcal{R}^n$

$$5) T(c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m) = c_1T(\mathbf{u}_1) + \dots + c_mT(\mathbf{u}_m)$$

The statements above for linear transformations are equivalent to the matrix statements where \mathbf{A} is a $m \times n$ matrix, \mathbf{u} and \mathbf{v} are vectors in \mathcal{R}^m and c is a scalar:

- 3) $\mathbf{A}\mathbf{0} = \mathbf{0}$
- 4) $\mathbf{A}(c\mathbf{u} + d\mathbf{v}) = c\mathbf{A}\mathbf{u} + d\mathbf{A}\mathbf{v}$

And for a $m \times n$ matrix \mathbf{A} , scalars c_1, \dots, c_m , and vectors $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathcal{R}^n$

$$5) \mathbf{A}(c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m) = c_1\mathbf{A}\mathbf{u}_1 + \dots + c_m\mathbf{A}\mathbf{u}_m$$

6.2 Types of matrix transformations

The basic types of matrix transformations include

- 1) contractions/expansions
- 2) rotations
- 3) reflections
- 4) shears
- 5) projections

For the following examples, we will consider the unit vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and apply different linear transformations using the matrix \mathbf{A} .

To build the matrix transformations, we use the `dasc2594` package and build matrix transformations based on code from <https://www.bryanshalloway.com/2020/02/20/visualizing-matrix-transformations-with-gganimate/>.

6.2.1 Contractions/Expansions

6.2.1.1 Horizontal Expansion

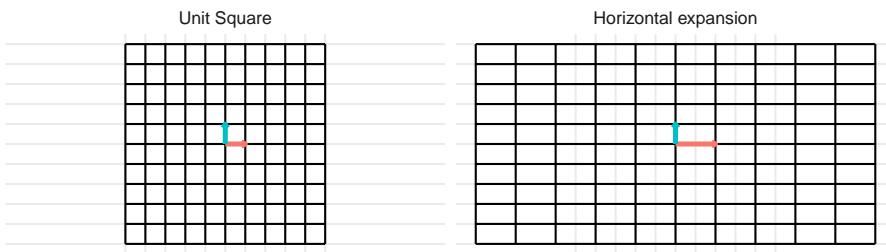
The matrix below gives a horizontal expansion when $x > 1$

$$\mathbf{A} = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix}$$

- In the example below, we set $x = 2$ and generate the transformation.

```
transformation_matrix <- tribble(
  ~ x, ~ y,
  2, 0,
  0, 1) %>%
  as.matrix()

p <- plot_transformation(transformation_matrix)
```

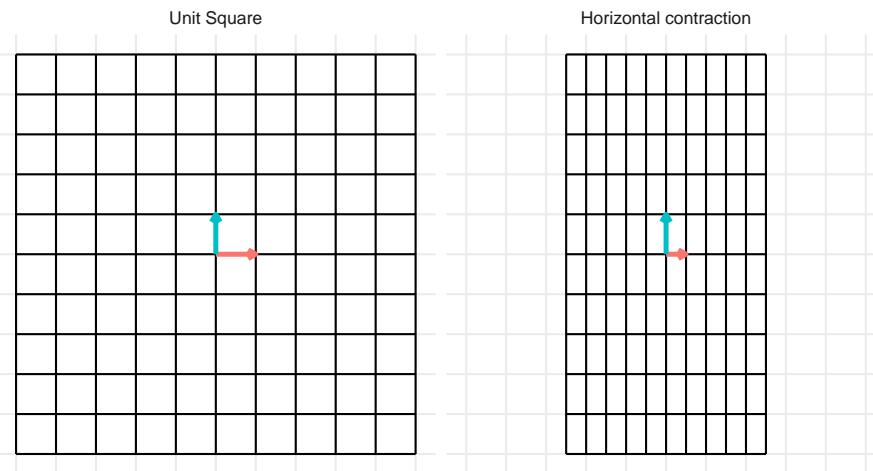


6.2.1.2 Horizontal Contraction

The matrix below gives a horizontal contraction when $x < 1$ * Horizontal contraction when $x < 1$

$$\mathbf{A} = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix}$$

- In the example below, we set $x = 0.5$

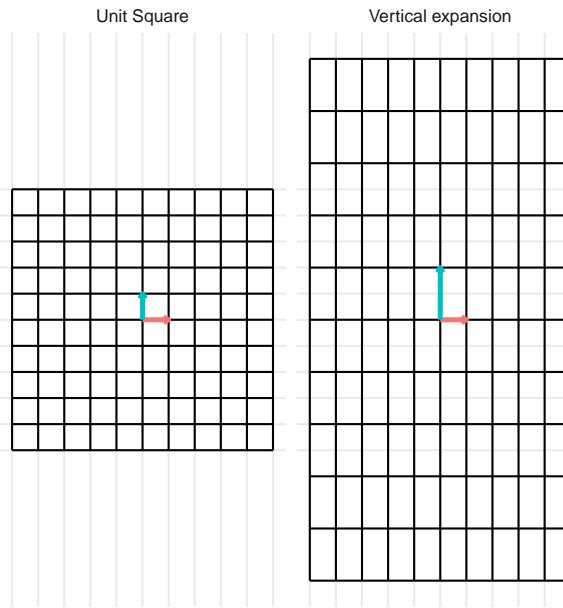


6.2.1.3 Vertical Expansion

The matrix below gives a vertical expansion when $x > 1$

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}$$

- * In the example below, we set $x = 2$

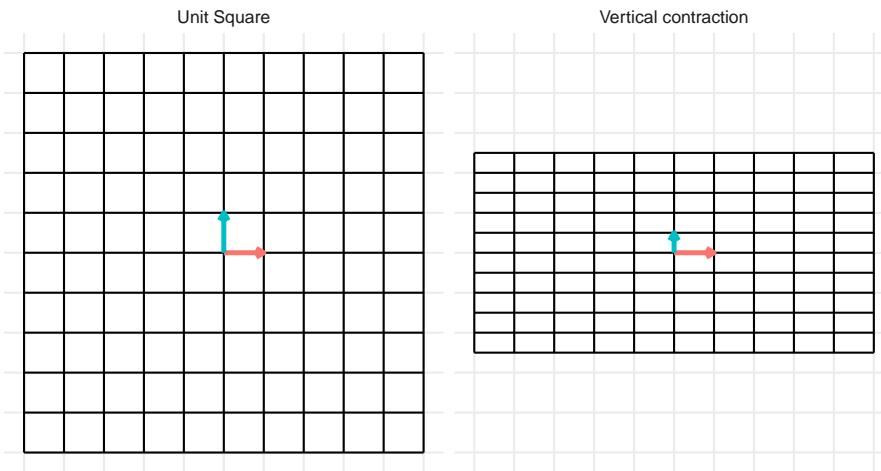


6.2.1.4 Vertical Contraction

The matrix below gives a vertical contraction when $x < 1$

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}$$

- In the example below, we set $x = 0.5$



6.2.2 Rotations

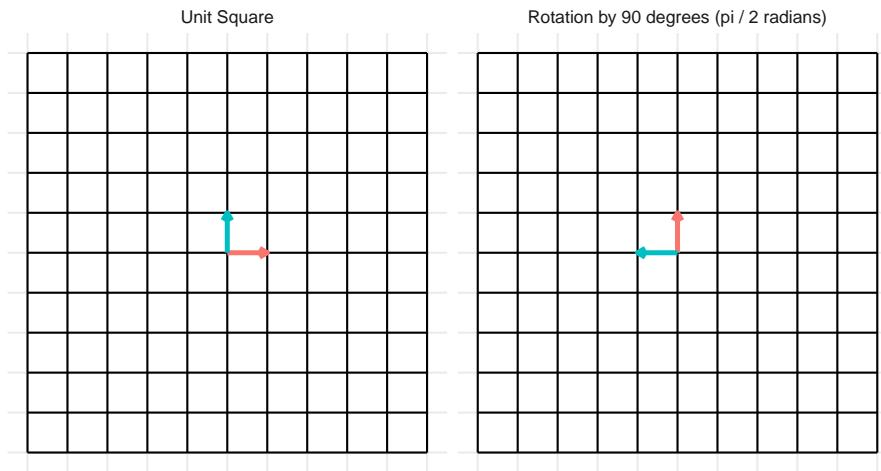
6.2.2.1 Rotation by 90 degrees

Rotations in 2D of an angle $\theta \in [0, 2\pi]$ take the form of

$$\mathbf{A} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

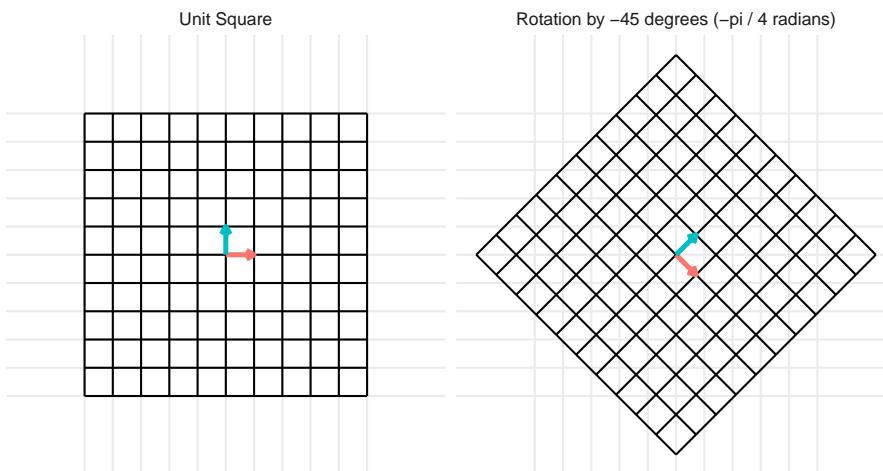
For example, a rotation of 90 degrees counter-clockwise ($\theta = \frac{\pi}{2}$) is given by the transformation matrix

$$\mathbf{A} = \begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$



Another example is for a rotation of 45 degrees clockwise ($\theta = -\frac{\pi}{4}$) is given by the transformation matrix

$$= \begin{pmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

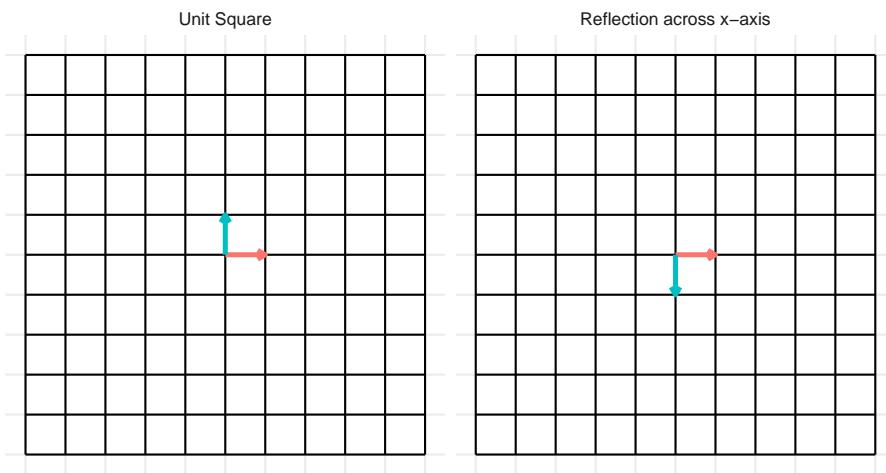


6.2.3 Reflections

6.2.3.1 Reflection across the x-axis

The matrix below gives a reflection about the x-axis

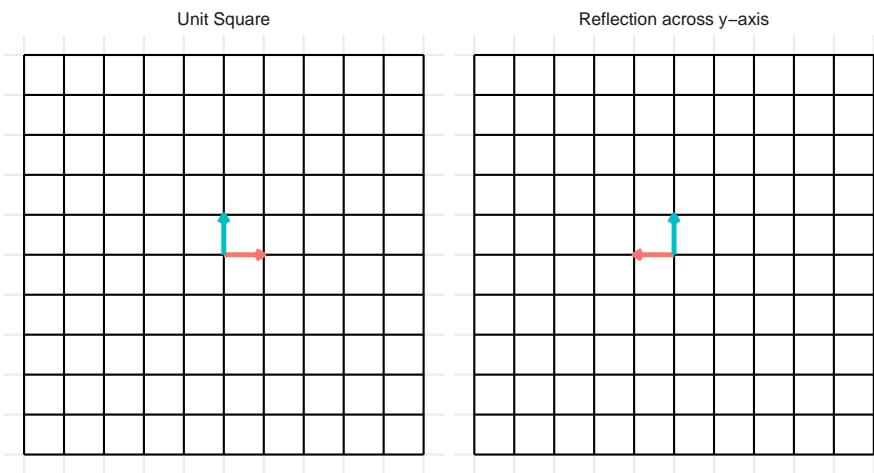
$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$



6.2.3.2 Reflection across the y-axis

The matrix below gives a reflection about the y-axis

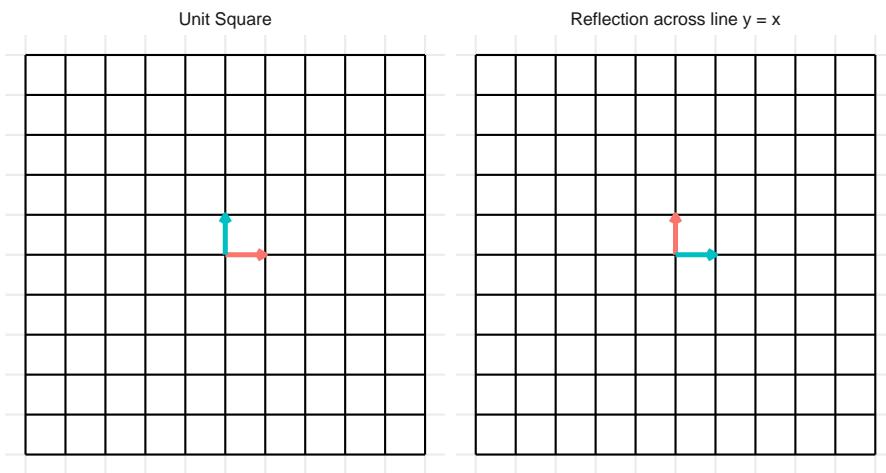
$$\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$



6.2.3.3 Reflection across the line $y = x$

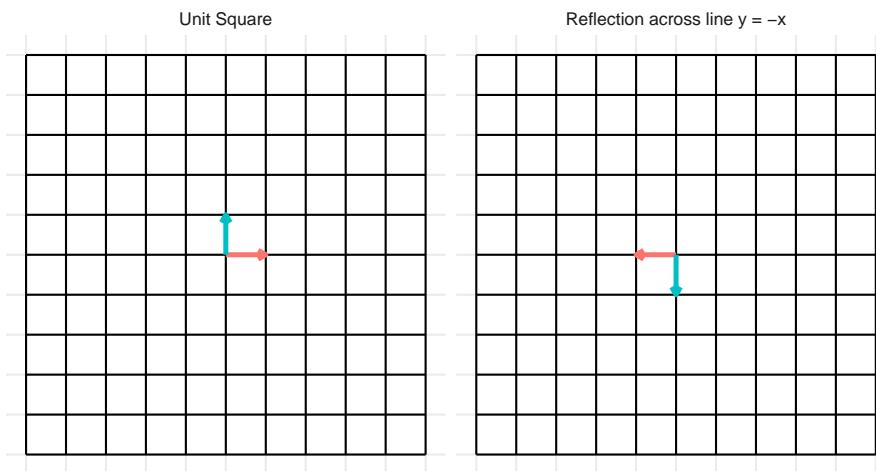
$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

- In the example below, we set $x = 0.5$



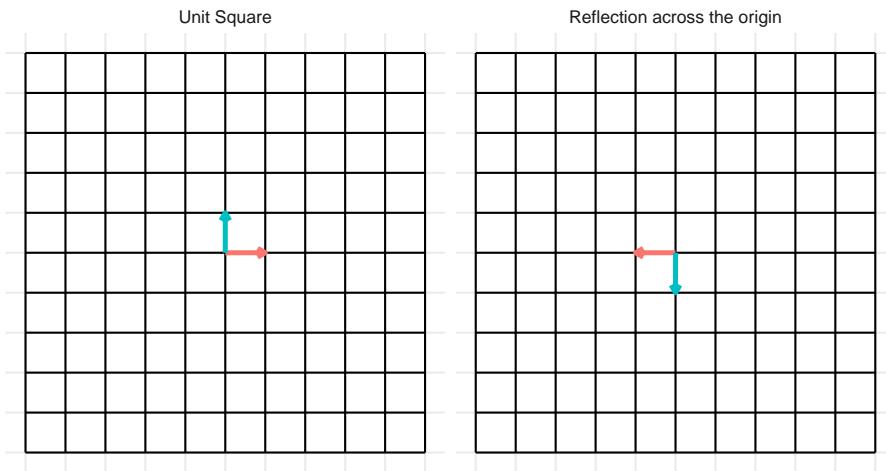
6.2.3.4 Reflection across the line $y = -x$

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$



6.2.3.5 Reflection across the origin (0, 0)

$$\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$



6.2.4 Shears

A shear transformation is like stretching play-dough if it was possible to stretch all parts of the dough uniformly (rather than some sections getting stretched more than others).

6.2.4.1 Horizontal Shear

$$\mathbf{A} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

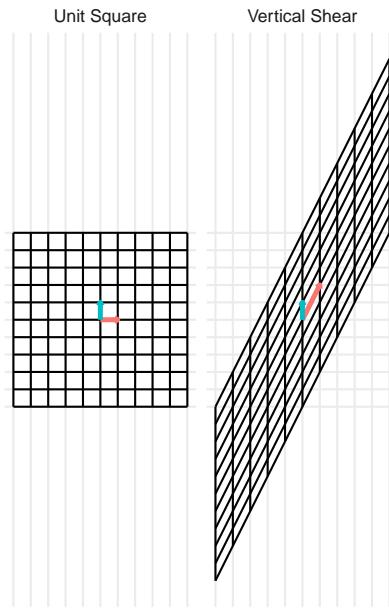
For the example below, we plot a horizontal shear with $x = 2$.



6.2.4.2 Vertical Shear

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}$$

For the example below, we plot a horizontal shear with $x = 2$.



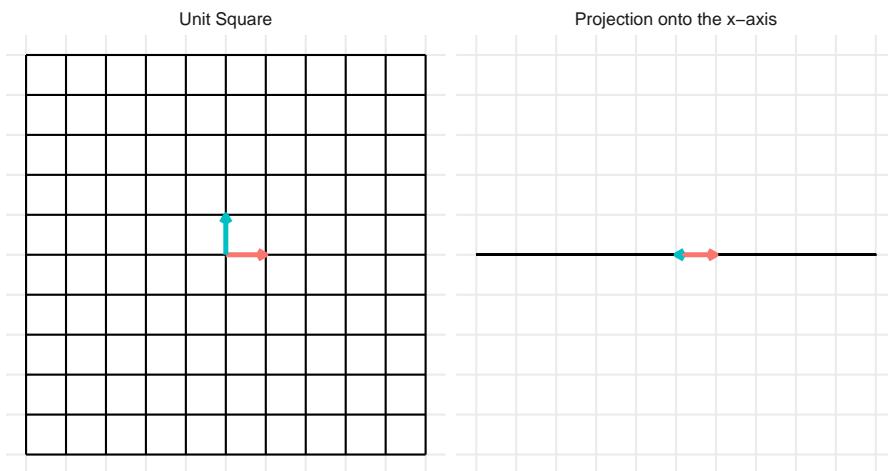
6.2.5 Projections

A projection is a mapping $T : \mathcal{R}^n \rightarrow \mathcal{R}^n$ from one space (\mathbf{R}^n) to itself (\mathbf{R}^n) such that $T^2 = T$

6.2.5.1 Project onto the x-axis

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

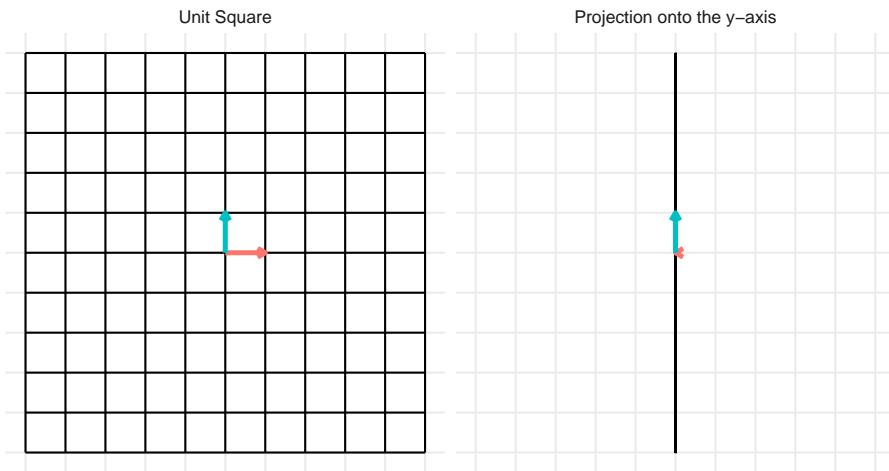
For the example below, we plot a projection onto the x-axis



6.2.5.2 Project onto the y-axis

$$\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

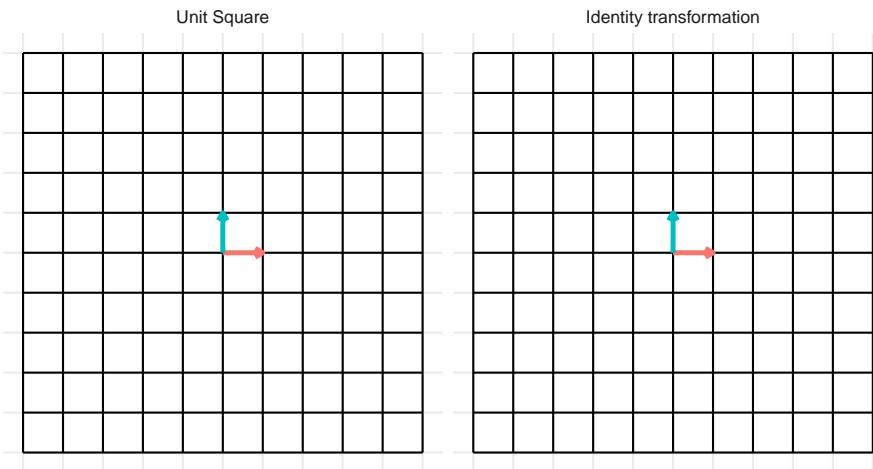
For the example below, we plot a projection onto the y-axis



6.2.6 Identity

The identity transformation is the transformation that leaves the vector input unchanged. The identity matrix is typically written as \mathbf{I}

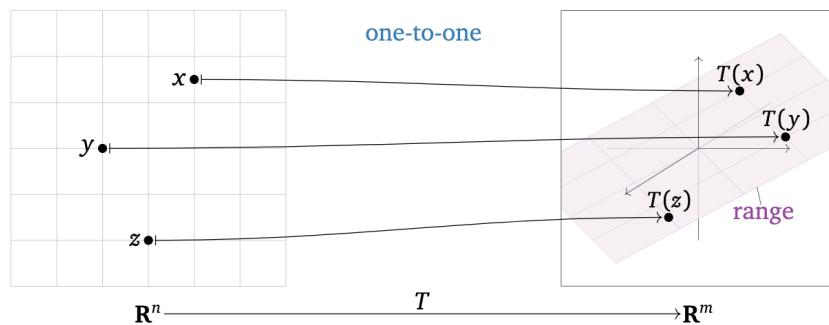
$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

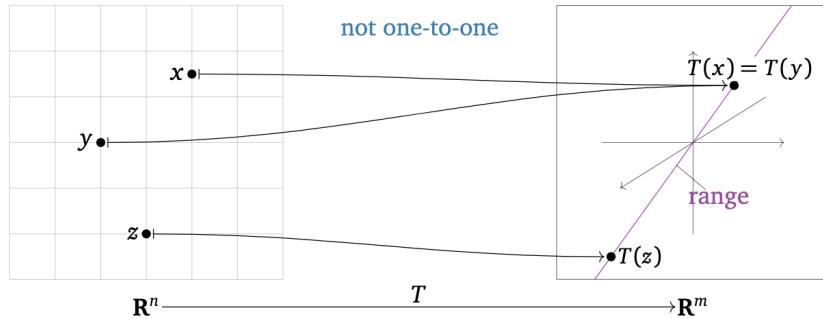


6.3 Properties of matrix transformations

6.3.1 One-to-one transformations

Definition 6.2. A transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called **one-to-one** if every vector \mathbf{b} in the image \mathbb{R}^m , the equation $T(\mathbf{x}) = \mathbf{b}$ has **at most** one solution in \mathbb{R}^n .





The following statements are equivalent was of saying $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is one-to-one:

- a) For every $\mathbf{b} \in \mathcal{R}^m$ (for every vector in the image), the equation $T(\mathbf{x}) = \mathbf{b}$ has either zero or one solution
- b) Every different input into the function $T(\cdot)$ has a different output
- c) If $T(\mathbf{u}) = T(\mathbf{v})$ then $\mathbf{u} = \mathbf{v}$

The following statements are equivalent was of saying $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is **not** one-to-one:

- a) There exists at least one $\mathbf{b} \in \mathcal{R}^m$ such that the equation $T(\mathbf{x}) = \mathbf{b}$ has more than one solution
- b) There are at least two different inputs into the function $T(\cdot)$ that have the same output
- c) There exist vectors $\mathbf{u} \neq \mathbf{v} \in \mathcal{R}^n$ such that $T(\mathbf{u}) = T(\mathbf{v})$

Theorem 6.1. Let \mathbf{Ax} be the matrix representation of the linear transformation $T(\mathbf{x})$ for the $m \times n$ matrix \mathbf{A} . Then the following statements are equivalent:

- 1) T is one-to-one.
- 2) For every $\mathbf{b} \in \mathcal{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at most one solution.
- 3) For every $\mathbf{b} \in \mathcal{R}^m$, the equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution or is inconsistent.
- 4) The equation $\mathbf{Ax} = \mathbf{0}$ has only a trivial solution.
- 5) The columns of \mathbf{A} are linearly independent.
- 6) \mathbf{A} has a pivot in every column.

7) The range of \mathbf{A} has dimension n

- **Example:** is the following matrix one-to-one?

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

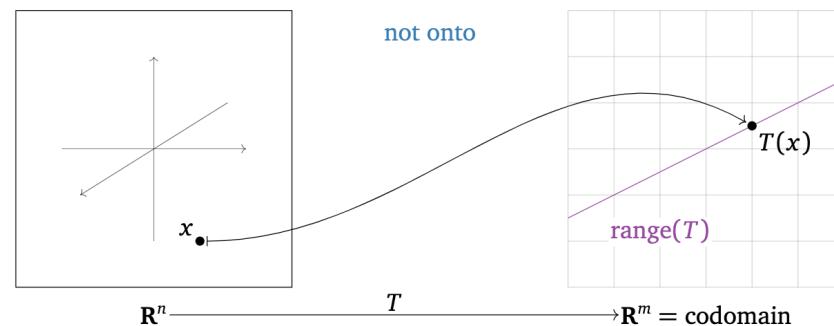
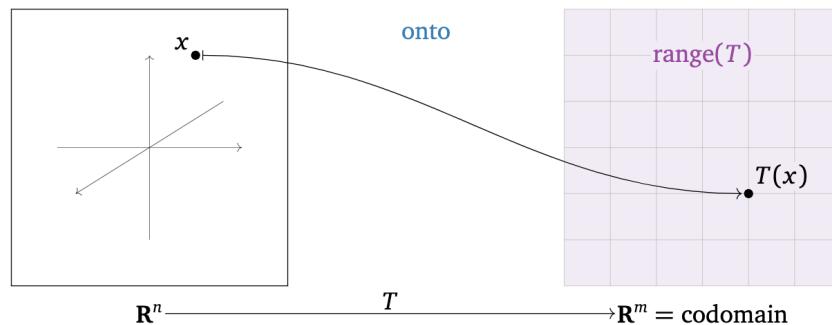
- **Example:** is the following matrix one-to-one?

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Note: Matrices that are wider than they are tall are not one-to-one transformations. (This does not mean that all tall matrices are one-to-one)

6.3.2 Onto transformations

Definition 6.3. A transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called **onto** if, for every vector $\mathbf{b} \in \mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution $\mathbf{x} \in \mathbb{R}^n$



The following are equivalent ways of saying that $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is onto:

- 1) The range of $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is equal to the codomain of $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$
- 2) Every vector in the codomain is the output of some input vector

The following are equivalent ways of saying that $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is not onto:

- 1) The range of $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ is smaller than the codomain of $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$.
- 2) There exists a vector $\mathbf{b} \in \mathcal{R}^m$ such that the equation $T(\mathbf{x})$ does not have a solution.
- 3) There is a vector in the codomain that is not the output of any input vector.

Theorem 6.2. *Let \mathbf{Ax} be the matrix representation of the linear transformation $T(\mathbf{x})$ for the $m \times n$ matrix \mathbf{A} . Then the following statements are equivalent:*

- 1) T is onto
- 2) $T(\mathbf{x}) = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathcal{R}^m$.
- 3) The equation $\mathbf{Ax} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathcal{R}^m$.
- 4) The columns of \mathbf{A} span \mathcal{R}^m
- 5) \mathbf{A} has a pivot in every row
- 6) The range of $T : \mathcal{R}^n \rightarrow \mathcal{R}^m$ has dimension m

- **Example:**

- **Example:** is the following matrix onto?

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

- **Example:** is the following matrix one-to-one?

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Note: Matrices that are taller than they are wide are not onto transformations.
(This does not mean that all wide matrices are onto)

Chapter 7

Matrix operations

7.1 ToDo

- define dot product here
- Show matrix multiplication using dot products
- move to chapter 1 or 2
- 3 Blue 1 Brown – Matrix Multiplication
- Note: add examples:

7.2 Properties of matrices

7.2.1 Matrix Addition

Matrix Addition: If the matrices \mathbf{A} and \mathbf{B} are of the same dimension (e.g., both \mathbf{A} and \mathbf{B} have the same number of rows m and the same number of columns n), then

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} + b_{11} & b_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \\ &= \{a_{ij} + b_{ij}\}\end{aligned}$$

... Another way to

If **A** and **B** are of the same dimension (same number of rows and columns) you can add the matrices together

```
A <- matrix(c(4, 1, 33, 2, 0, -4), 3, 2)
B <- matrix(c(7, -24, 3, 9, 11, -9), 3, 2)
A
```

```
##      [,1] [,2]
## [1,]     4     2
## [2,]     1     0
## [3,]    33    -4
```

```
B
```

```
##      [,1] [,2]
## [1,]     7     9
## [2,]   -24    11
## [3,]     3    -9
```

```
A + B
```

```
##      [,1] [,2]
## [1,]    11    11
## [2,]   -23    11
## [3,]    36   -13
```

We can also write this using **for** loops

```
# initialize an empty matrix to fill
C <- matrix(0, 3, 2)

for (i in 1:nrow(A)) {          # loop over the rows
  for (j in 1:ncol(A)) {        # loop over the columns
    C[i, j] <- A[i, j] + B[i, j]
  }
}
C
```

```
##      [,1] [,2]
## [1,]    11    11
## [2,]   -23    11
## [3,]    36   -13
```

If \mathbf{A} and \mathbf{B} are of different dimensions (they differ in either the number of rows or columns), R will return an error warning you that the matrices are of different sizes and can't be added

```
A <- matrix(c(4, 1, 33, 2, 0, -4), 3, 2)
B <- matrix(c(7, -24, 3, 9), 2, 2)
A

##      [,1] [,2]
## [1,]     4    2
## [2,]     1    0
## [3,]    33   -4

B

##      [,1] [,2]
## [1,]     7    3
## [2,]   -24    9

A + B

## Error in A + B: non-conformable arrays
```

Theorem 7.1. Let \mathbf{A} , \mathbf{B} , and \mathbf{C} be $m \times n$ matrices and let a and b be scalars, then:

- 1) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ [(commutivity of addition)]
- 2) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ [(commutivity of addition)]
- 3) $\mathbf{A} + \mathbf{0} = \mathbf{A}$ [(additive identity)]
- 4) $a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B}$ [(scalar ...)]
- 5) $(a + b)\mathbf{A} = a\mathbf{A} + b\mathbf{A}$ [(scalar ...)]
- 6) $(ab)\mathbf{A} = a(b\mathbf{A})$ [(scalar ...)]

7.2.2 Matrix Multipliation

Matrix Multiplication: If $\mathbf{A} = \{a_{ij}\}$ is an $m \times n$ matrix and $\mathbf{B} = \{a_{jk}\}$ is a $n \times p$ matrix, then the matrix product $\mathbf{C} = \mathbf{AB}$ is an $m \times p$ matrix where $\mathbf{C} = \left\{\sum_{j=1}^p a_{ij}b_{jk}\right\}$

$$\begin{aligned}\mathbf{AB} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^n a_{1j}b_{j1} & \sum_{j=1}^n a_{1j}b_{j2} & \cdots & \sum_{j=1}^n a_{1j}b_{jp} \\ \sum_{j=1}^n a_{2j}b_{j1} & \sum_{j=1}^n a_{2j}b_{j2} & \cdots & \sum_{j=1}^n a_{2j}b_{jp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n a_{mj}b_{j1} & \sum_{j=1}^n a_{mj}b_{j2} & \cdots & \sum_{j=1}^n a_{mj}b_{jp} \end{pmatrix}\end{aligned}$$

Another way to define matrix multiplication is through inner product notation. Define the $m \times n$ matrix \mathbf{A} and the $n \times p$ matrix \mathbf{B} as the partition

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_m \end{pmatrix} \quad \text{and} \quad \mathbf{B} = (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p)$$

where \mathbf{a}_i and \mathbf{b}_k are both n -vectors. Then, we have $\mathbf{C} = \mathbf{AB}$ can be written as

$$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p) = (\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \cdots \quad \mathbf{Ab}_p)$$

Note that in this representation, each column of the matrix \mathbf{AB} is a linear combination the the columns of \mathbf{A} with coefficients given by the corresponding column of \mathbf{B} .

$$\begin{aligned}\mathbf{AB} &= \begin{pmatrix} \mathbf{a}'_1\mathbf{b}_1 & \mathbf{a}'_1\mathbf{b}_2 & \cdots & \mathbf{a}'_1\mathbf{b}_q \\ \mathbf{a}'_2\mathbf{b}_1 & \mathbf{a}'_2\mathbf{b}_2 & \cdots & \mathbf{a}'_2\mathbf{b}_q \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}'_n\mathbf{b}_1 & \mathbf{a}'_n\mathbf{b}_2 & \cdots & \mathbf{a}'_n\mathbf{b}_q \end{pmatrix} \\ &= \{\mathbf{a}'_i\mathbf{b}_k\}.\end{aligned}$$

Written in this notation, we arrive at the multiplication rule for $\mathbf{C} = \mathbf{AB}$ – the ik th element c_{ik} of \mathbf{C} is the inner product of the i th row of \mathbf{A} and the j th column of \mathbf{B} .

7.2.3 Properties of Matrix Multiplication

Define the $m \times m$ **identity matrix** \mathbf{I}_m with ones on the diagonal and zeros off diagonal as

$$\mathbf{I}_m = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Let \mathbf{A} be an $m \times n$ matrix, then:

- 1) Let \mathbf{B} be an $n \times p$ matrix and \mathbf{C} a $p \times q$ matrix. Then $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ is an $m \times q$ matrix.
- 2) Let \mathbf{B} and \mathbf{C} be $n \times p$ matrices. Then $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ is an $m \times p$ matrix.
- 3) Let \mathbf{B} and \mathbf{C} be $p \times m$ matrices. Then $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$ is an $p \times m$ matrix.
- 4) Let \mathbf{B} be an $p \times m$ matrix and c a scalar. Then $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$ is an $p \times m$ matrix.
- 5) $\mathbf{I}_m \mathbf{A} = \mathbf{AI}_n = \mathbf{A}$

Examples: in class

Note: Matrix multiplication violates some of the rules of multiplication that you might be used to. Pay attention for the following:

- 1) In general $\mathbf{AB} \neq \mathbf{BA}$ (sometimes these are equal, but usually are not)
- 2) $\mathbf{AB} = \mathbf{AC}$ does not imply $\mathbf{B} = \mathbf{C}$
- 3) $\mathbf{AB} = \mathbf{0}$ does not imply that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$

7.2.4 Matrix Multiplication complexity (Big O notation)

In the study of algorithms, the notation $O(n)$ is used to describe the number of calculations that need to be done to evaluate the equation. As an example, consider $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 2 & -3 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix}$, and $\mathbf{x} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$.

By hand: Calculate

- 1) $(\mathbf{AB})\mathbf{x}$
- 2) $\mathbf{A}(\mathbf{Bx})$

Which was easier? Which required less calculation?

- Matrix-matrix multiplication of an $m \times n$ matrix \mathbf{A} and an $m \times p$ matrix \mathbf{B} has complexity $O(mnp)$.
- Matrix-vector multiplication of an $m \times n$ matrix \mathbf{A} and an p -vector \mathbf{x} has complexity $O(nm)$.

From example above:

- 1) $O(mnp)$ matrix-matrix multiplication (\mathbf{AB}) followed by $O(mn)$ matrix-vector multiplication $(\mathbf{AB})\mathbf{x}$ which has computational complexity $O(mnp) + O(mn)$
- 2) $O(mn)$ matrix-vector multiplication (\mathbf{Bx}) followed by $O(mn)$ matrix-vector multiplication $\mathbf{A}(\mathbf{Bx})$ which has computational complexity $O(mn) + O(mn)$

7.2.5 Matrix powers

Powers of a $n \times n$ (square) matrix are defined as the product of \mathbf{A} multiplied k times

$$\mathbf{A}^k = \underbrace{\mathbf{A} \cdots \mathbf{A}}_k$$

7.2.6 Matrix Transpose

The matrix transpose is an operator that swaps the rows and columns of a matrix. If \mathbf{A} is an $m \times n$ matrix (m rows and n columns), then \mathbf{A}' is a $n \times m$ matrix (n rows and m columns). Note: some use \mathbf{A}^T to denote a transpose; I prefer the ' notation as it is much simpler and cleaner notation). The matrix

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

has transpose

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix},$$

Theorem 7.2. Let \mathbf{A} be an $m \times n$ matrix, then

- 1) $(\mathbf{A}')' = \mathbf{A}$.
- 2) Let \mathbf{B} be an $m \times n$ matrix, then $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.
- 3) For any scalar c , $(c\mathbf{A})' = c\mathbf{A}'$.
- 4) Let \mathbf{B} be an $n \times p$ matrix, then $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ is an $p \times m$ matrix.

Note: The power of video games: GPUs and modern CPUs are becoming more and more parallelized. Because the ij th element of \mathbf{AB} requires only the i th row of \mathbf{A} and the j th column of \mathbf{B} , matrix multiplication is easily parallelized under modern computing architectures. Thanks to video games, this parallelization has been made faster than ever.

Examples: in class

Chapter 8

Matrix Inverses

- 3 Blue 1 Brown – Inverse Matrices, column space, and null space

```
library(dasc2594)
library(tidyverse)
```

For scalars, the multiplicative identity is

$$a \frac{1}{a} = aa^{-1} = a^{-1}a = 1$$

where a^{-1} is the inverse of a .

Definition 8.1 (Matrix Inverse). The $n \times n$ **square** matrix \mathbf{A} is said to be **invertible** if there exists a $n \times n$ matrix \mathbf{C} (which we call \mathbf{A}^{-1} once we verify the inverse exists) such that

$$\begin{aligned}\mathbf{CA} &= \mathbf{AC} = \mathbf{I} \\ \mathbf{A}^{-1}\mathbf{A} &= \mathbf{AA}^{-1} = \mathbf{I}\end{aligned}$$

where \mathbf{I} is the $n \times n$ identity matrix (the matrix with 1s on the diagonal and zeros everywhere else).

In R, an identity matrix is easy to construct. An $n \times n$ identity matrix can be constructed using the `diag()` function

```
n <- 4
I <- diag(n)
I
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
```

Example 8.1.

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 2 & -3 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 3 & -1 \\ 2 & -1 \end{pmatrix}$$

```
# check if B is the inverse of A
A %*% B
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

```
# check if B is the inverse of A
B %*% A
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

Because $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, we have \mathbf{A} is an invertible matrix with inverse $\mathbf{B} = \mathbf{A}^{-1}$.

Theorem 8.1. Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $ad - bc \neq 0$ then \mathbf{A} is invertible and

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If $ad - bc = 0$, then the matrix is not invertible.

- **Question:** why is the matrix not invertible when $ad - bc = 0$?

- Have you heard of “singular” or “singularity” before?
- Black holes are called singularities. Why is this?
- Square matrices that are not invertible are call “singular”

Definition 8.2. For the 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the term $ad - bc$ is called the **determinant** of the matrix \mathbf{A} and is written as $\det(\mathbf{A})$. Sometimes the determinant is written as $|\mathbf{A}|$

A consequence of the above theorem is that a 2×2 matrix is invertible only if its determinant is nonzero.

Example 8.2. Determine if the following 2×2 matrix is invertible

$$\mathbf{A} = \begin{pmatrix} 4 & -4 \\ -1 & 2 \end{pmatrix}$$

Theorem 8.2. If the $n \times n$ matrix \mathbf{A} is invertible, then for each $\mathbf{b} \in \mathcal{R}^n$, the matrix equation

$$\mathbf{Ax} = \mathbf{b}$$

has the unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Proof. There are two things to show ...

- 1) show there is a solution
- 2) show the solution is unique

□

Example 8.3. Let $\mathbf{A} = \begin{pmatrix} 4 & -4 & -2 \\ 5 & 2 & -5 \\ -4 & 6 & 1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

Find the solution to $\mathbf{Ax} = \mathbf{b}$

Theorem 8.3 (Invertible Matrix Theorem). 1) If \mathbf{A} is an invertible matrix, then \mathbf{A}^{-1} is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

2) If \mathbf{A} and \mathbf{B} are $n \times n$ invertible matrices, then \mathbf{AB} is also an invertible matrix whose inverse is

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

which is the inverse of the matrices in reverse order.

3) If \mathbf{A} is an invertible matrix, then the transpose \mathbf{A}' is also invertible and the inverse of \mathbf{A}' is the transpose of \mathbf{A}^{-1} . Equivalently,

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

Proof. Here we prove the three statements from the theorem above. All three statements rely on the definition of an invertible matrix in Definition 8.1

1) If \mathbf{A}^{-1} is invertible, then, there exists a matrix \mathbf{C} such that $\mathbf{CA}^{-1} = \mathbf{A}^{-1}\mathbf{C} = \mathbf{I}$. Let $\mathbf{C} = \mathbf{A}$. Then, we have $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ which shows that $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

- 2) First, consider multiplying \mathbf{AB} on the left by $\mathbf{B}^{-1}\mathbf{A}^{-1}$ where $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{A}\mathbf{I}\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$. Then multiply \mathbf{AB} on the right by $\mathbf{B}^{-1}\mathbf{A}^{-1}$ where $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}(\mathbf{AA}^{-1})\mathbf{B}^{-1} = \mathbf{B}\mathbf{I}\mathbf{B}^{-1} = \mathbf{BB}^{-1} = \mathbf{I}$.
- 3) Use the fact that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. Then, $(\mathbf{A}^{-1})'\mathbf{A}' = (\mathbf{AA}^{-1})' = \mathbf{I}' = \mathbf{I}$. Similarly $\mathbf{A}'(\mathbf{A}^{-1})' = (\mathbf{A}^{-1}\mathbf{A})' = \mathbf{I}' = \mathbf{I}$. Thus \mathbf{A}' is invertible with inverse $(\mathbf{A}^{-1})'$

□

- **Note:** A consequence of theorem 8.3 (2) is that the product of k invertible $n \times n$ matrices $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_k$ has inverse $\mathbf{A}_k^{-1}\mathbf{A}_{k-1}^{-1} \cdots \mathbf{A}_1^{-1}$

8.1 Elementary matrices

- Elementary matrices are matrices that perform basic row operations (i.e., we can write the reduced row echelon algorithm as a product of elementary matrices).

Recall the elementary row operations:

- 1) swaps: swapping two rows.
- 2) sums: replacing a row by the sum itself and a multiple of another row.
- 3) scalar multiplication: replacing a row by a scalar multiple times itself.

- **Example:** Consider a 3×3 matrix

```
A <- matrix(c(4, 5, 9, -2, -4, 1, 4, 6, -2), 3, 3)
```

$$\mathbf{A} = \begin{pmatrix} 4 & -2 & 4 \\ 5 & -4 & 6 \\ 9 & 1 & -2 \end{pmatrix}$$

- 1) What is the elementary matrix (let's call it \mathbf{E}_1) that swaps the first and second rows of \mathbf{A} ?

```
E_1 <- matrix(c(0, 1, 0, 1, 0, 0, 0, 0, 1), 3, 3)
```

$$\mathbf{E}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A

```
##      [,1] [,2] [,3]
## [1,]     4   -2    4
## [2,]     5   -4    6
## [3,]     9     1   -2

## left multiple A by E_1
E_1 %*% A
```

```
##      [,1] [,2] [,3]
## [1,]     5   -4    6
## [2,]     4   -2    4
## [3,]     9     1   -2
```

Thus, the matrix $\mathbf{E}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is the matrix that swaps the first and second row.

- 2) What is the elementary matrix (let's call it \mathbf{E}_2) that adds two times the first of \mathbf{A} to the third row of \mathbf{A} ?

```
E_2 <- matrix(c(1, 0, 2, 0, 1, 0, 0, 0, 1), 3, 3)
```

$$\mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

A

```
##      [,1] [,2] [,3]
## [1,]     4   -2    4
## [2,]     5   -4    6
## [3,]     9     1   -2

## left multiple A by E_2
E_2 %*% A
```

```
##      [,1] [,2] [,3]
## [1,]     4   -2    4
## [2,]     5   -4    6
## [3,]    17   -3    6
```

Thus, the matrix $\mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$ is the matrix that adds two times the first of \mathbf{A} to the third row of \mathbf{A}

- 3) What is the elementary matrix (let's call it \mathbf{E}_3) that multiplies the second row of \mathbf{A} by 3?

```
E_3 <- matrix(c(1, 0, 0, 0, 3, 0, 0, 0, 1), 3, 3)
```

$$\mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A

```
##      [,1] [,2] [,3]
## [1,]     4   -2    4
## [2,]     5   -4    6
## [3,]     9     1   -2

## left multiple A by E_3
E_3 %*% A
```

```
##      [,1] [,2] [,3]
## [1,]     4   -2    4
## [2,]    15  -12   18
## [3,]     9     1   -2
```

Thus, the matrix $\mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is the matrix that multiplies the second row of \mathbf{A} by 3.

- **Question:** Do you see any patterns with how the example elementary matrices look?

$$\mathbf{E}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad \mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- The elementary matrices look like the identity matrix \mathbf{I} with an elementary row operation applied to \mathbf{I} . In fact, this leads us to this general fact:

Fact: If an elementary row matrix is applied to the $m \times n$ matrix \mathbf{A} , the result of this elementary row operation applied to \mathbf{A} can be written as \mathbf{EA} where \mathbf{E} is the $m \times m$ identity matrix \mathbf{I} with the respective elementary row operation applied to \mathbf{I} .

Fact: Each elementary matrix \mathbf{E} is invertible

Example: in class

The next theorem is quite important as the result gives an algorithm for calculating the inverse of a $n \times n$ matrix \mathbf{A} which also makes it possible to solve matrix equations $\mathbf{Ax} = \mathbf{b}$

Theorem 8.4. *If an $n \times n$ matrix \mathbf{A} is invertible, then \mathbf{A} is row-equivalent to \mathbf{I} ($\mathbf{A} \sim \mathbf{I}$; row-equivalent means \mathbf{A} can be reduced to \mathbf{I} using elementary row operations). The row-equivlency implies that there is a series of elementary row operations (e.g., elementary matrices $\mathbf{E}_1, \dots, \mathbf{E}_k$) that converts \mathbf{A} to \mathbf{I} . In addition, the application of these row matrices to \mathbf{I} transforms \mathbf{I} to the matrix inverse \mathbf{A}^{-1} .*

- **Proof:** in class

8.2 Finding the inverse of A

The previous theorem states that for a $n \times n$ invertible matrix \mathbf{A} , the elementary row operations that convert \mathbf{A} to \mathbf{I} also convert \mathbf{I} to \mathbf{A}^{-1} . This suggests an algorithm for finding the inverse \mathbf{A}^{-1} of \mathbf{A} :

Create the augmented matrix $(\mathbf{A} \mid \mathbf{I})$ and row reduce the augmented matrix. If the row-reduced augmented matrix is of the form $(\mathbf{I} \mid \mathbf{A}^{-1})$ then \mathbf{A}^{-1} is the inverse of \mathbf{A} . If the leading matrix in the augmented matrix is not the identity matrix \mathbf{I} , then \mathbf{A} is not row equivalent to \mathbf{I} and is therefore not invertible.

Example 8.4. Let $\mathbf{A} = \begin{pmatrix} -3 & -3 & -4 \\ -4 & 2 & -4 \\ 4 & -4 & 4 \end{pmatrix}$. Does \mathbf{A} have an inverse, and if so, what is it?

Using R

8.3 The Invertible Matrix Theorem

Theorem 8.5 (The Invertible Matrix Theorem). *Let \mathbf{A} be an $n \times n$ matrix. Then the following statements are equivalent (i.e., they are all either simultaneously true or false).*

- 1) \mathbf{A} is an invertible matrix.
- 2) \mathbf{A} is row equivalent to the $n \times n$ identity matrix \mathbf{I} ($\mathbf{A} \sim \mathbf{I}$).
- 3) \mathbf{A} and n pivot columns.
- 4) The homogeneous matrix equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
- 5) The columns of \mathbf{A} are linearly independent.
- 6) The linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ given by the matrix transformation $\mathbf{x} \rightarrow \mathbf{Ax}$ is one-to-one.
- 7) The inhomogeneous matrix equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{R}^n$.
- 8) The columns of \mathbf{A} span \mathbb{R}^n .
- 9) The linear transformation $\mathbf{x} \rightarrow \mathbf{Ax}$ maps \mathbb{R}^n onto \mathbb{R}^n .
- 10) There is an $n \times n$ matrix \mathbf{C} such that $\mathbf{CA} = \mathbf{I}$.
- 11) There is an $n \times n$ matrix \mathbf{D} such that $\mathbf{AD} = \mathbf{I}$.
- 12) \mathbf{A}' is an invertible matrix.

Proof. In class □

A result of the invertible matrix theorem is that if \mathbf{A} and \mathbf{B} are $n \times n$ matrices with $\mathbf{AB} = \mathbf{I}$ then $\mathbf{A} = \mathbf{B}^{-1}$ and $\mathbf{B} = \mathbf{A}^{-1}$.

8.4 Invertible Linear Transformations

Definition 8.3. A linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be invertible if there exists a transformation $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n \quad T(S(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n$$

- Draw figure in class

Theorem 8.6. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation and let \mathbf{A} be the matrix representing the transformation T . Then the transformation T is invertible if and only if the matrix \mathbf{A} is invertible. Therefore, the matrix that represents $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the inverse transformation of T , is unique and is represented by the matrix \mathbf{A}^{-1} .

Chapter 9

Block Matrices

Another way to represent matrices is using a block (or partitioned) form. A block-representation of a matrix arises when the $n \times p$ matrix \mathbf{A} is represented using smaller blocks as follows:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1K} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{J1} & \mathbf{A}_{J2} & \cdots & \mathbf{A}_{JK} \end{pmatrix}$$

where \mathbf{A}_{ij} is a $n_j \times p_k$ matrix where $\sum_{j=1}^J n_j = n$ and $\sum_{k=1}^K p_k = p$.

For example, the matrix

$$\mathbf{A} = \begin{pmatrix} 5 & 7 & 1 \\ 5 & -22 & 2 \\ -14 & 5 & 99 \\ 42 & -3 & 0 \end{pmatrix},$$

can be written in block matrix form with

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \begin{bmatrix} 5 & 7 \\ 5 & -22 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ \begin{bmatrix} -14 & 5 \\ 42 & -3 \end{bmatrix} & \begin{bmatrix} 99 \\ 0 \end{bmatrix} \end{pmatrix}, \end{aligned}$$

where $\mathbf{A}_{11} = \begin{bmatrix} 5 & 7 \\ 5 & -22 \end{bmatrix}$ is a 2×2 matrix, $\mathbf{A}_{12} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is a 2×1 matrix, etc.

```

A_11 <- matrix(c(5, 5, 7, -22), 2, 2)
A_12 <- c(1, 2)
A_21 <- matrix(c(-14, 42, 5, -3), 2, 2)
A_22 <- c(99, 0)

## bind columns then rows
rbind(
  cbind(A_11, A_12),
  cbind(A_21, A_22)
)

##          A_12
## [1,]    5    7    1
## [2,]    5   -22   2
## [3,]   -14    5   99
## [4,]    42   -3    0

## bind rows then columns
cbind(
  rbind(A_11, A_21),
  c(A_12, A_22) ## rbind on vectors is different than c()
)

##      [,1] [,2] [,3]
## [1,]    5    7    1
## [2,]    5   -22   2
## [3,]   -14    5   99
## [4,]    42   -3    0

## bind rows then columns
cbind(
  rbind(A_11, A_21),
  ## convert the vectors to matrices for rbind
  rbind(as.matrix(A_12), as.matrix(A_22))
)

##      [,1] [,2] [,3]
## [1,]    5    7    1
## [2,]    5   -22   2
## [3,]   -14    5   99
## [4,]    42   -3    0

```

9.1 Block Matrix Addition

If \mathbf{A} and \mathbf{B} are both $m \times n$ block matrices with blocks in r rows and c columns where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1c} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{r1} & \mathbf{B}_{r2} & \cdots & \mathbf{B}_{rc} \end{pmatrix}$$

and each block \mathbf{A}_{ij} and \mathbf{B}_{ij} have the same dimension, then

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} & \cdots & \mathbf{A}_{1c} + \mathbf{B}_{1c} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} & \cdots & \mathbf{A}_{2c} + \mathbf{B}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{r1} + \mathbf{B}_{r1} & \mathbf{A}_{r2} + \mathbf{B}_{r2} & \cdots & \mathbf{A}_{rc} + \mathbf{B}_{rc} \end{pmatrix}$$

which is a matrix where each block is the sum of the other blocks. Notice that if each block was a scalar rather than a block matrix, this would be the usual definition of matrix addition (compare equation (9.1) above to (7.2.1)). The one requirement is that each of the blocks \mathbf{A}_{ij} and \mathbf{B}_{ij} have the same dimension. When this is true, we say that \mathbf{A} and \mathbf{B} are **conformable for block matrix addition**.

9.2 Block Matrix Multiplication

If \mathbf{A} and \mathbf{B} are both $m \times n$ block matrices with blocks in r rows and c columns (same as above) where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1c} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{r1} & \mathbf{B}_{r2} & \cdots & \mathbf{B}_{rc} \end{pmatrix}$$

and each row of blocks \mathbf{A}_{ij} has the same number of columns as the block \mathbf{B}_{ij} has rows, then the block matrices \mathbf{A} and \mathbf{B} are said to be **conformable for block matrix multiplication**. A consequence of this is that $r = c$. When this is the case, the matrix products is

$$\mathbf{AB} = \begin{pmatrix} \sum_{j=1}^c \mathbf{A}_{1j} \mathbf{B}_{j1} & \sum_{j=1}^c \mathbf{A}_{1j} \mathbf{B}_{j2} & \cdots & \sum_{j=1}^c \mathbf{A}_{1j} \mathbf{B}_{jc} \\ \sum_{j=1}^c \mathbf{A}_{2j} \mathbf{B}_{j1} & \sum_{j=1}^c \mathbf{A}_{2j} \mathbf{B}_{j2} & \cdots & \sum_{j=1}^c \mathbf{A}_{2j} \mathbf{B}_{jc} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^c \mathbf{A}_{rj} \mathbf{B}_{j1} & \sum_{j=1}^c \mathbf{A}_{rj} \mathbf{B}_{j2} & \cdots & \sum_{j=1}^c \mathbf{A}_{rj} \mathbf{B}_{jc} \end{pmatrix}$$

which can be said in words as "each block-element (the ij th element $(\mathbf{AB})_{ij}$) of the block-matrix product \mathbf{AB} is the sum of the i th block-row of \mathbf{A} and the j th block column of \mathbf{B} . Notice that if each block was a scalar rather than a block matrix, this would be the usual definition of matrix multiplication (compare equation (9.1) above to (7.2.2)).

Example 9.1. in class

Solution. A solution to the example problem

9.3 The column-row matrix product

Theorem 9.1. *The matrix product \mathbf{AB} of an $m \times n$ matrix $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n)$ with columns $\{\mathbf{a}_i\}_{i=1}^n$ and an $n \times p$ matrix $\mathbf{B} = \begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_n \end{pmatrix}$ with rows $\{\mathbf{b}'_i\}_{i=1}^n$ can be written as the column-row expansion below:*

$$\begin{aligned} \mathbf{AB} &= (\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n) \begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_n \end{pmatrix} \\ &= \mathbf{a}_1 \mathbf{b}'_1 + \mathbf{a}_2 \mathbf{b}'_2 + \cdots + \mathbf{a}_n \mathbf{b}'_n \end{aligned}$$

Recall: The notation \mathbf{b}'_i has a transpose because a vector is defined in the vertical orientation (column vector). Therefore, to formally define a row vector, we take a vertical vector of the values in the row and take its transpose to turn the column vector into a row vector.

Example 9.2. in class

9.4 Special Block Matrices

There are many different forms of block matrices. Two that deserve special mention here include block diagonal matrices and block triangular matrices.

Definition 9.1. The matrix \mathbf{A} is said to be block diagonal if

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_n \end{pmatrix}$$

Definition 9.2. The matrix \mathbf{A} is said to be block (upper) triangular if

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \cdots & \mathbf{A}_{1n} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{A}_{23} & \cdots & \mathbf{A}_{2n} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{33} & \cdots & \mathbf{A}_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{nn} \end{pmatrix}$$

\mathbf{A} is block (lower) triangular if

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \mathbf{A}_{m3} & \cdots & \mathbf{A}_{mn} \end{pmatrix}$$

Example 9.3. Assume that \mathbf{A} , which has the form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix},$$

is an invertible matrix where \mathbf{A}_{11} a $p \times p$ invertible matrix, \mathbf{A}_{12} a $p \times q$ matrix, and \mathbf{A}_{22} is a $q \times q$ invertible matrix. Solve for \mathbf{A}^{-1}

Solution. The inverse to \mathbf{A} is a matrix $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$ with blocks \mathbf{B}_{11} , \mathbf{B}_{12} , \mathbf{B}_{21} , and \mathbf{B}_{22} of appropriate size to be conformable with the blocks \mathbf{A}_{11} , \mathbf{A}_{12} , and \mathbf{A}_{22} that make up the matrix \mathbf{A} .

For \mathbf{B} to be the inverse of \mathbf{A} , $\mathbf{AB} = \mathbf{BA} = \mathbf{I} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$ where $\mathbf{0}$ is a matrix of zeros of the appropriate size. Writing this out in blocks gives

$$\begin{aligned} \mathbf{BA} &= \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{B}_{11}\mathbf{A}_{11} & \mathbf{B}_{11}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22} \\ \mathbf{B}_{21}\mathbf{A}_{11} & \mathbf{B}_{21}\mathbf{A}_{12} + \mathbf{B}_{22}\mathbf{A}_{22} \end{pmatrix} \end{aligned}$$

which gives that $\mathbf{B}_{11} = \mathbf{A}_{11}^{-1}$ because $\mathbf{B}_{11}^{-1}\mathbf{A}_{11} = \mathbf{I}$ and \mathbf{A}_{11} is invertible. The equation also give $\mathbf{B}_{21}\mathbf{A}_{11} = \mathbf{0}$ and because \mathbf{A}_{11} is an invertible matrix, the homogeneous equation $\mathbf{A}_{11}\mathbf{b} = \mathbf{0}$ has only the trivial solution for each column \mathbf{b} of the matrix \mathbf{B}_{21} which implies that $\mathbf{B}_{21} = \mathbf{0}$. Using these facts, we can rewrite the above equation as

$$\mathbf{BA} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22} \\ \mathbf{0} & \mathbf{B}_{22}\mathbf{A}_{22} \end{pmatrix}$$

Because the lower right entry $\mathbf{B}_{22}\mathbf{A}_{22}$ must equal \mathbf{I} , we have that $\mathbf{B}_{22} = \mathbf{A}_{22}^{-1}$. Then, the equation becomes

$$\mathbf{BA} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

The final component is the upper right block. Because we are finding the inverse, we know that $\mathbf{0} = \mathbf{A}_{11}^{-1}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22}$. Subtracting $\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ from both sides of the equation gives

$$\begin{aligned} \mathbf{B}_{12}\mathbf{A}_{22} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{B}_{12}\mathbf{A}_{22}\mathbf{A}_{22}^{-1} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{B}_{12}\mathbf{I} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{aligned}$$

Thus, $\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix}$

Chapter 10

Matrix Factorizations

```
library(tidyverse)
library(dasc2594)
library(mvtnfast)
library(MASS)
```

In scalar mathematics, a factorization is an expression that writes a scalar a as a product of two or more scalars. For example, the scalar 2 has a square-root factorization of $2 = \sqrt{2} * \sqrt{2}$ and 15 has a prime factorization of $15 = 3 * 5$. A matrix factorization is a similar concept where a matrix \mathbf{A} can be represented by a product of two or more matrices (e.g., $\mathbf{A} = \mathbf{BC}$). In data science, matrix factorizations are fundamental to working with data.

10.1 The LU factorization

First, we define lower and upper triangular matrices.

Definition 10.1. The matrix \mathbf{A} is said to be lower triangular if

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

Similarly, the matrix \mathbf{A} is said to be upper triangular if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

The LU factorization of a matrix \mathbf{A} reduces the matrix \mathbf{A} into two components. The first component \mathbf{L} is a lower-triangular matrix and the second component \mathbf{U} is an upper triangular matrix.

Using the LU factorization, the matrix factorization $\mathbf{A} = \mathbf{LU}$ can be used in the matrix equation $\mathbf{Ax} = \mathbf{LUx} = \mathbf{b}$ by first solving the sub-equation $\mathbf{Ly} = \mathbf{b}$ and then solving the second sub-equation $\mathbf{Ux} = \mathbf{y}$ for \mathbf{x} . Thus, the matrix factorization applied to the matrix equation gives the pair of equations

$$\begin{aligned} \mathbf{Ly} &= \mathbf{b} \\ \mathbf{Ux} &= \mathbf{y} \end{aligned}$$

At first glance, this seems like we are trading the challenge of solving one system of equations \mathbf{Ax} (4) for the two equations in (10.1). However, the computational benefits arise due to the fact that \mathbf{L} and \mathbf{U} are triangular matrices and solving matrix equations with triangular matrices is much faster.

Example 10.1. Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & -2 \\ -2 & -2 & -4 & 1 \\ -1 & -4 & -8 & 5 \\ -2 & -6 & -4 & 4 \end{pmatrix}$ which has the LU decomposition

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & -2 \\ -2 & -2 & -4 & 1 \\ -1 & -4 & -8 & 5 \\ -2 & -6 & -4 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 \\ -1 & -2 & -3 & 0 \\ -2 & -3 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & -2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

and consider the system of equations defined by the matrix equation $\mathbf{Ax} = \mathbf{b}$

$$\text{where } \mathbf{b} = \begin{pmatrix} -5 \\ -7 \\ -2 \\ -14 \end{pmatrix}.$$

- 1) solve $\mathbf{Ly} = \mathbf{b}$ using an augmented matrix and RREF.
- 2) solve $\mathbf{Ux} = \mathbf{y}$ using an augmented matrix and RREF.

- 3) compare to the solution $\mathbf{Ax} = \mathbf{b}$ using an augmented matrix and RREF.

Solution. For the example, we will show how to solve a system of equations using the LU decomposition for the equation defined above.

- 1) Solve $\mathbf{Ly} = \mathbf{b}$ using augmented matrix

$$\begin{array}{l}
 \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ -2 & -1 & 0 & 0 & -7 \\ -1 & -2 & -3 & 0 & -2 \\ -2 & -3 & 0 & -3 & -14 \end{array} \right) \xrightarrow[r_2 \leftarrow -\frac{1}{2}r_2 - r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1/2 & 0 & 0 & 17/2 \\ -1 & -2 & -3 & 0 & -2 \\ -2 & -3 & 0 & -3 & -14 \end{array} \right) \\
 \xrightarrow[r_3 \leftarrow -r_3 - r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1/2 & 0 & 0 & 17/2 \\ 0 & 2 & 3 & 0 & 7 \\ -2 & -3 & 0 & -3 & -14 \end{array} \right) \xrightarrow[r_4 \leftarrow -\frac{1}{2}r_4 - r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1/2 & 0 & 0 & 17/2 \\ 0 & 2 & 3 & 0 & 7 \\ 0 & 3/2 & 0 & 3/2 & 12 \end{array} \right) \\
 \xrightarrow[r_2 \leftarrow 2r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 & 17 \\ 0 & 2 & 3 & 0 & 7 \\ 0 & 3/2 & 0 & 3/2 & 12 \end{array} \right) \xrightarrow[r_3 \leftarrow \frac{1}{2}r_3 - r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 & 17 \\ 0 & 0 & 3/2 & 0 & -27/2 \\ 0 & 3/2 & 0 & 3/2 & 12 \end{array} \right) \\
 \xrightarrow[r_4 \leftarrow \frac{2}{3}r_4 - r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 & 17 \\ 0 & 0 & 3/2 & 0 & -27/2 \\ 0 & 0 & 0 & 1 & -9 \end{array} \right) \xrightarrow[r_3 \leftarrow \frac{2}{3}r_3]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 & 17 \\ 0 & 0 & 1 & 0 & -9 \\ 0 & 0 & 0 & 1 & -9 \end{array} \right)
 \end{array}$$

- 2) solve $\mathbf{Ux} = \mathbf{y}$ using an augmented matrix and RREF.

$$\begin{array}{l}
 \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 2 & 0 & 3 & 17 \\ 0 & 0 & 2 & -3 & -9 \\ 0 & 0 & 0 & -3 & -9 \end{array} \right) \xrightarrow[r_2 \leftarrow \frac{1}{2}r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 2 & -3 & -9 \\ 0 & 0 & 0 & -3 & -9 \end{array} \right) \\
 \xrightarrow[r_3 \leftarrow \frac{1}{2}r_3]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & -3 & -9 \end{array} \right) \xrightarrow[r_4 \leftarrow -\frac{1}{3}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \\
 \xrightarrow[r_1 \leftarrow r_1 - 2r_3]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 1 & 4 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \xrightarrow[r_1 \leftarrow r_1 - 2r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \\
 \xrightarrow[r_2 \leftarrow r_2 - \frac{3}{2}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \xrightarrow[r_3 \leftarrow r_3 + \frac{3}{2}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right)
 \end{array}$$

3) compare to the solution $\mathbf{Ax} = \mathbf{b}$ using an augmented matrix and RREF.

$$\begin{array}{c}
 \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ -2 & -2 & -4 & 1 & -7 \\ -1 & -4 & -8 & 5 & -2 \\ -2 & -6 & -4 & 4 & -14 \end{array} \right) \xrightarrow[r_2 \leftarrow r_2 + 2r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & -2 & 0 & -3 & -17 \\ -1 & -4 & -8 & 5 & -2 \\ -2 & -6 & -4 & 4 & -14 \end{array} \right) \\
 \xrightarrow[r_3 \leftarrow r_3 + r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & -2 & 0 & -3 & -17 \\ 0 & -4 & -6 & 3 & -7 \\ -2 & -6 & -4 & 4 & -14 \end{array} \right) \xrightarrow[r_4 \leftarrow r_4 + 2r_1]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & -2 & 0 & -3 & -17 \\ 0 & -4 & -6 & 3 & -7 \\ 0 & -6 & 0 & 0 & -24 \end{array} \right) \\
 \xrightarrow[r_2 \leftarrow -\frac{1}{2}r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & -4 & -6 & 3 & -7 \\ 0 & -6 & 0 & 0 & -24 \end{array} \right) \xrightarrow[r_3 \leftarrow r_3 + 4r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & -6 & 9 & 27 \\ 0 & -6 & 0 & 0 & -24 \end{array} \right) \\
 \xrightarrow[r_4 \leftarrow r_4 + 6r_2]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & -6 & 9 & 27 \\ 0 & 0 & 0 & 9 & 27 \end{array} \right) \xrightarrow[r_3 \leftarrow -\frac{1}{6}r_3]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 9 & 27 \end{array} \right) \\
 \xrightarrow[r_4 \leftarrow \frac{1}{9}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 2 & -2 & -5 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \xrightarrow[r_1 \leftarrow r_1 - 2r_3]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 1 & 4 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \\
 \xrightarrow[r_1 \leftarrow r_1 - r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3/2 & 17/2 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \xrightarrow[r_2 \leftarrow r_2 - \frac{3}{2}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & -3/2 & -9/2 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right) \\
 \xrightarrow[r_3 \leftarrow r_3 + \frac{3}{2}r_4]{\sim} \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right)
 \end{array}$$

While it might not be completely obvious, once one has calculated the LU decomposition, it can often be much faster to solve systems of equations with the LU decomposition.

Exercise 10.1. in lab: Solve some large systems of equations by brute force which shows how the LU decomposition is faster.

10.1.1 Geometric interpretation of the LU factorization

- Draw image in class – composition of transformations $T_A(\cdot) = T_L(T_U(\cdot))$

10.2 Obtaining the LU factorization

Notice that the upper-triangular matrix \mathbf{U} is in echelon form. Congratulations! you know how to construct a matrix \mathbf{U} by reducing the matrix \mathbf{A} to an echelon form \mathbf{U} using elementary matrices $\mathbf{E}_1, \dots, \mathbf{E}_k$. Now, we only need to find the lower triangular matrix \mathbf{L} .

Combining the LU factorization and the fact that we can find an upper triangular matrix \mathbf{U} using elementary row matrices, we have

$$\begin{aligned}\mathbf{A} &= \mathbf{LU} \\ \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} &= \mathbf{U}.\end{aligned}$$

We also know that each of the elementary row matrices \mathbf{E}_j are invertible (you can always re-swap rows, subtract instead of add rows, etc.) which says that each inverse \mathbf{E}_j^{-1} exists. Thus, the product $\mathbf{E}_k \cdots \mathbf{E}_1$ must have an inverse which is

$$(\mathbf{E}_k \cdots \mathbf{E}_1)^{-1} = \mathbf{E}_1^{-1} \cdots \mathbf{E}_k^{-1}.$$

Plugging this inverse into (10.2) gives (left multiplying by $(\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}$ on both sides)

$$\begin{aligned}(\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}(\mathbf{E}_k \cdots \mathbf{E}_1)\mathbf{A} &= (\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}\mathbf{U} \\ \mathbf{A} &= (\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}\mathbf{U} \\ &= \mathbf{LU}\end{aligned}$$

where $\mathbf{L} = (\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}$

Algorithm for finding the LU decomposition

Given the matrix \mathbf{A}

- 1) Find elementary matrices $\mathbf{E}_1, \dots, \mathbf{E}_k$ such that $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A}$ is in row echelon form (if this is possible, otherwise an LU factorization does not exist). Call this matrix \mathbf{U} , the upper triangular component of the LU factorization.
- 2) The, the lower triangular $\mathbf{L} = (\mathbf{E}_k \cdots \mathbf{E}_1)^{-1}$.

Notice that the algorithm does not say to find a specific matrix \mathbf{U} . In general, any row echelon form matrix \mathbf{U} will work.

10.3 The Cholesky factor

A Cholesky decomposition is special type of LU decomposition. A Cholesky decomposition is an LU decomposition on a **symmetric, positive-definite** square matrix.

Definition 10.2. • A matrix \mathbf{A} is said to be symmetric if $\mathbf{A} = \mathbf{A}'$

- A $n \times n$ matrix is said to be positive definite if for all $\mathbf{x} \in \mathcal{R}^n$, the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$

Note: the condition of positive definiteness is actually impossible to check. Can you show this is true for all vectors? Luckily, a $n \times n$ symmetric matrix is positive definite if and only if the matrix \mathbf{A} is invertible (which we know about by the invertible matrix theorem 8.5).

Definition 10.3. Let \mathbf{A} be a symmetric, positive definite matrix (by this, \mathbf{A} is a $n \times n$ square matrix). Then

$$\mathbf{A} = \mathbf{L}\mathbf{L}'$$

is the Cholesky decomposition of \mathbf{A} if \mathbf{L} is a lower-triangular matrix. Also, the lower triangular Cholesky matrix \mathbf{L} is unique.

What makes the Cholesky factor special?

- The decomposition $\mathbf{A} = \mathbf{L}\mathbf{U}$ has the property that $\mathbf{U} = \mathbf{L}'$ so that the computer only has to store one of the matrix components (reduce memory demands). As about half of the elements of \mathbf{L} are 0, matrix multiplication is much less computationally demanding as about half of the flops are not required to be evaluated ($x * 0 = 0$).
- The Cholesky factor is unique. There is only one Cholesky factor for each symmetric positive definite matrix.
- The Cholesky has properties related to multivariate normal distributions.

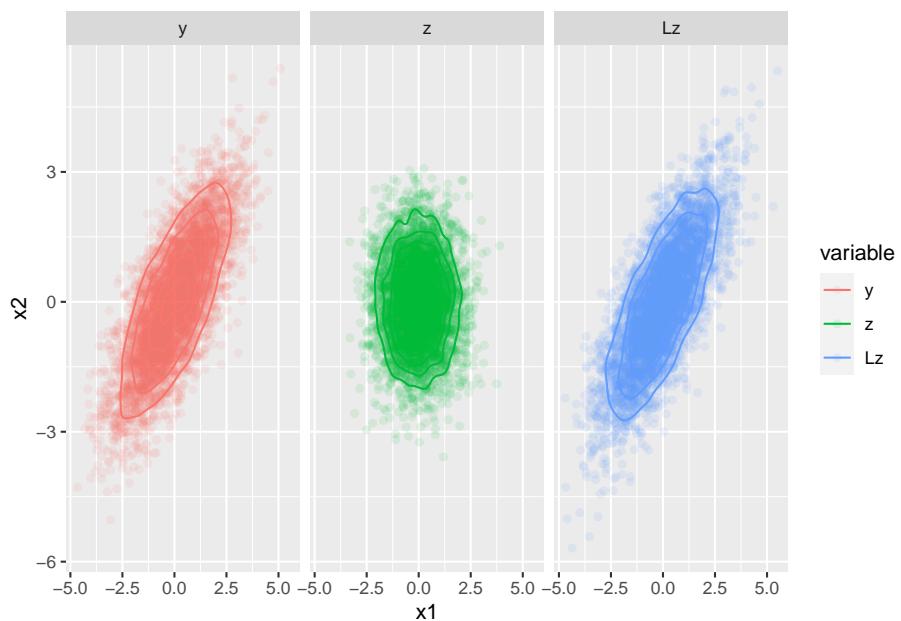
Let $\mathbf{y} \sim N(\mathbf{0}, \Sigma)$, and $\Sigma = \mathbf{L}\mathbf{L}'$. Then, if $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$, then $\mathbf{L}\mathbf{z} \sim N(\mathbf{0}, \Sigma)$. We say the \mathbf{y} and $\mathbf{L}\mathbf{z}$ are equal in distribution.

```
# simulate N 2-dimensional random normal vectors
N <- 5000
mu     <- rep(0, 2)
Sigma <- matrix(c(2, 1.5, 1.5, 2), 2, 2)
y <- rmvnorm(N, mu, Sigma)

# calculate the Cholesky factor
L <- t(chol(Sigma))      # R calculates the upper (right) Cholesky factor by default
z <- rmvnorm(N, mu, diag(2))
Lz <- t(L %*% t(z))    # pay attention to the dimensions of L and z here...

data.frame(
  observation = 1:N,
```

```
x1      = c(y[, 1], z[, 1], Lz[, 1]),
x2      = c(y[, 2], z[, 2], Lz[, 2]),
variable = factor(rep(c("y", "z", "Lz"), each = N), levels = c("y", "z", "Lz"))
) %>%
  ggplot(aes(x = x1, y = x2, color = variable)) +
  geom_point(alpha = 0.1) +
  geom_density2d() +
  facet_wrap(~ variable)
```



Chapter 11

Subspaces of \mathcal{R}^n

- 3 Blue 1 Brown – Linear combinations, span, and basis vectors
- 3 Blue 1 Brown – Inverse Matrices, column space, and null space

```
library(dasc2594)
```

First, let's recall the definition of a subset. A set A is a subset of a set B if all elements of A are also members of B . For example, the integers \mathbb{Z} are a subset of the real numbers \mathbf{R} ($\mathbb{Z} \subset \mathbf{R}$) and the real numbers are a subset of the complex numbers \mathcal{C} ($\mathbf{R} \subset \mathcal{C}$).

Subspaces are a generalization of the idea of subsets that are useful for understanding vector spaces.

Definition 11.1. A subspace \mathcal{H} of \mathcal{R}^n is a set that has the properties

- 1) The zero vector $\mathbf{0} \in \mathcal{H}$ [additive identity]
- 2) For each $\mathbf{u}, \mathbf{v} \in \mathcal{H}$, the sum $\mathbf{u} + \mathbf{v}$ is in \mathcal{H} [closed under vector addition]
- 3) For each $\mathbf{u} \in \mathcal{H}$ and scalar c , the scalar multiple $c\mathbf{u}$ is in \mathcal{H} [closed under scalar multiplication]

Example 11.1. Let \mathbf{u} and \mathbf{v} be vectors in \mathcal{R}^n . Then the vector space defined by $\text{span}\{\mathbf{u}, \mathbf{v}\}$ is a subspace of \mathcal{R}^n

Solution. To show that $\text{span}\{\mathbf{u}, \mathbf{v}\}$ is a subspace of \mathcal{R}^n , we must satisfy the three conditions of definition (11.1).

- 1) $\mathbf{0} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$ because $\mathbf{0} = 0\mathbf{u} + 0\mathbf{v}$ is a linear combination of \mathbf{u} and \mathbf{v} with coefficients 0 and 0.

- 2) Let $\mathbf{x} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$ and $\mathbf{y} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$. Then, by the definition of the span (Definition (3.1)) there exists constants a, b, c, d such that $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$ and $\mathbf{y} = c\mathbf{u} + d\mathbf{v}$. Thus, $\mathbf{x} + \mathbf{y} = a\mathbf{u} + b\mathbf{v} + c\mathbf{u} + d\mathbf{v} = (a+c)\mathbf{u} + (b+d)\mathbf{v}$ which, by definition, is in the $\text{span}\{\mathbf{u}, \mathbf{v}\}$. Thus, $\text{span}\{\mathbf{u}, \mathbf{v}\}$ is closed under addition.
- 3) Let $\mathbf{x} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$ c be a constant. Because $\mathbf{x} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$, the definition of the span (Definition (3.1)) implies there exists constants a and b such that $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$. Thus, $c\mathbf{x} = ca\mathbf{u} + cb\mathbf{v}$ which is a linear combination of \mathbf{u} and \mathbf{v} and is therefore $c\mathbf{x} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$.

Because all three requirements are met, we have shown that $\text{span}\{\mathbf{u}, \mathbf{v}\}$ is a subspace of \mathcal{R}^n .

Example 11.2. Recall that a solution to the matrix equation $\mathbf{Ax} = \mathbf{0}$ that has one free variable is a line through the origin and the solution to the matrix equation $\mathbf{Ax} = \mathbf{b}$ that has one free variable is a line parallel to the prior line that does not go through the origin.

- A) Is a line through the origin a subspace?
- B) Is a line **not** through the origin a subspace?

Solution. **Part A:** First, we will consider whether the line through the origin is a subspace. To show this, we need to verify the three properties of a subspace in Definition (11.1)

- 1) $\mathbf{0}$ is on the line through the origin because $\mathbf{Ax} = \mathbf{0}$ has the trivial solution $\mathbf{x} = \mathbf{0}$.
- 2) If \mathbf{u} and \mathbf{v} are on the line through the origin, then because the line through the origin is defined as the solution set of the equation $\mathbf{Ax} = \mathbf{0}$ we know that

$$\mathbf{Au} = \mathbf{0} \quad \text{and} \quad \mathbf{Av} = \mathbf{0}.$$

Then, the vector $\mathbf{u} + \mathbf{v}$ is in the subspace defined by the line through the origin because

$$\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

which is a solution to the matrix equation and therefore $\mathbf{u} + \mathbf{v}$ is a point on the line through the origin.

- 3) If \mathbf{u} is a point on the line through the origin and c is a scalar, then because the line through the origin is defined as the solution set of the equation $\mathbf{Ax} = \mathbf{0}$ we know that

$$\mathbf{Au} = \mathbf{0}.$$

Then, the vector $c\mathbf{u}$ is in the subspace defined by the line through the origin because

$$\mathbf{A}(c\mathbf{u}) = c\mathbf{A}\mathbf{u} = c\mathbf{0} = \mathbf{0},$$

which is a solution to the matrix equation and therefore $c\mathbf{u}$ is a point on the line through the origin.

Part B: Next, we consider whether the line **not** through the origin is a subspace. To show this, we need to verify the three properties of a subspace in Definition (11.1). Recall that a line not through the origin is a solution set of the matrix equation $\mathbf{Ax} = \mathbf{b}$ where the solution set has a single free variable and $\neq \mathbf{0}$. To check if this line defines a subspace, we check the first condition of Definition (11.1)

- 1) $\mathbf{0}$ is not on the line through the origin because $\mathbf{Ax} = \mathbf{b}$ does not have the trivial solution $\mathbf{x} = \mathbf{0}$ because $\mathbf{A}\mathbf{0} = \mathbf{0} \neq \mathbf{b}$. Because $\mathbf{0}$ is not a point on the line not through the origin, the line not through the origin does not define a subspace.
- **Note:** For any vectors $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^n$, the $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a subspace of \mathbb{R}^n .

11.1 Special subspaces: column space and null space

Definition 11.2. The **column space**, denoted $\text{col}(\mathbf{A})$, of a $m \times n$ matrix \mathbf{A} which has columns $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ is the set of vectors that are linear combinations of the columns of \mathbf{A} which is equivalent to the $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

Example 11.3. Given the matrix $\mathbf{A} = \begin{pmatrix} 1 & -8 & -9 \\ -3 & 3 & -5 \\ 1 & 0 & 6 \\ -2 & -1 & -1 \end{pmatrix}$ with columns $\mathbf{a}_1, \mathbf{a}_2$,

and \mathbf{a}_3 , what is the column space of \mathbf{A} ?

Solution. Given the matrix $\mathbf{A} = \begin{pmatrix} 1 & -8 & -9 \\ -3 & 3 & -5 \\ 1 & 0 & 6 \\ -2 & -1 & -1 \end{pmatrix}$ with columns $\mathbf{a}_1, \mathbf{a}_2$, and \mathbf{a}_3 ,

the column space of \mathbf{A} is the set of vectors \mathbf{b} such that

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3.$$

Thus, the column space of \mathbf{A} is the span of the vectors that makes up the columns of \mathbf{A} .

Definition 11.3. The **null space**, denoted $\text{null}(\mathbf{A})$, of a matrix \mathbf{A} is the set of all solutions to the homogeneous matrix equation $\mathbf{Ax} = \mathbf{0}$.

While the idea of a null space might at first glance seem unclear, the null space is the set of all vectors which the matrix transformation defined by \mathbf{A} maps to $\mathbf{0}$. In other words, the null space of \mathbf{A} is the set of vectors $\{\mathbf{x} : \mathbf{Ax} = \mathbf{0}\}$.

Example 11.4. Given the matrix $\mathbf{A} = \begin{pmatrix} -3 & -3 & -4 & -5 & -2 \\ -4 & 2 & -4 & 5 & 3 \\ 4 & -4 & 4 & -3 & 5 \end{pmatrix}$ with columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and \mathbf{a}_4 , find vectors that span the null space of \mathbf{A} .

Solution. Given the matrix $\mathbf{A} = \begin{pmatrix} -3 & -3 & -4 & -5 & -2 \\ -4 & 2 & -4 & 5 & 3 \\ 4 & -4 & 4 & -3 & 5 \end{pmatrix}$ with columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and \mathbf{a}_4 , the null space of \mathbf{A} is the solution set of the homogeneous system of equations $\mathbf{Ax} = \mathbf{0}$. Using an augmented matrix and transforming into reduced row echelon form gives the RREF form

$$\begin{pmatrix} -3 & -3 & -4 & -5 & -2 & 0 \\ -4 & 2 & -4 & 5 & 3 & 0 \\ 4 & -4 & 4 & -3 & 5 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & -15 & -25 & 0 \\ 0 & 1 & 0 & -1 & -4 & 0 \\ 0 & 0 & 1 & 53/4 & 89/4 & 0 \end{pmatrix}$$

which tells us that the solution to the system of equations is

$$\begin{aligned} x_1 - 15x_4 - 25x_5 &= 0 \\ x_2 - x_4 - 4x_5 &= 0 \\ x_3 + \frac{53}{4}x_4 + \frac{89}{4}x_5 &= 0 \\ x_4 &= x_4 \\ x_5 &= x_5 \end{aligned}$$

Writing this solution as a vector times x_4 and a vector times x_5 gives the vectors

$$\begin{pmatrix} 15 \\ 1 \\ -\frac{53}{4} \\ 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 25 \\ 4 \\ -\frac{89}{4} \\ 0 \\ 1 \end{pmatrix}.$$

In R, this is shown by starting with the matrix \mathbf{A}

```
A <- matrix(c(-3, -4, 4, -3, 2, -4, -4, -4, 4, -5, 5, -3, -2, 3, 5), 3, 5)
A

##      [,1] [,2] [,3] [,4] [,5]
## [1,]    -3   -3   -4   -5   -2
## [2,]    -4    2   -4    5    3
## [3,]     4   -4    4   -3    5
```

Looking at the reduced row echelon form of \mathbf{A} gives

```
rref(A)

##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    0    0 -15.00 -25.00
## [2,]    0    1    0 -1.00  -4.00
## [3,]    0    0    1 13.25  22.25
```

where the columns of interest are the non-pivot columns. For this matrix \mathbf{A} , the fourth and fifth columns of \mathbf{A} are the non-pivot columns. The fourth column or the RREF form corresponds to the variable x_4 and the fifth column corresponds to the variable x_5 . Thus, you can extract the vectors that form the null space from these fourth and fifth columns of the RREF form of \mathbf{A} like so

```
# the nullspace vector corresponding to x4 = 1 and x5 = 0
c(-rref(A)[, 4], 1, 0)
```

```
## [1] 15.00 1.00 -13.25 1.00 0.00
```

```
# the nullspace vector corresponding to x4 = 0 and x5 = 1
c(-rref(A)[, 5], 0, 1)
```

```
## [1] 25.00 4.00 -22.25 0.00 1.00
```

We check that these vectors are in $\text{null}(\mathbf{A})$ by using matrix multiplication and verifying that these are zero (at least up to numeric overflow/underflow)

```
A %*% c(-rref(A)[, 4], 1, 0)
```

```
##      [,1]
## [1,] -7.105427e-15
## [2,] -7.105427e-15
## [3,]  7.105427e-15
```

```
A %*% c(-rref(A)[, 5], 0, 1)
```

```
##      [,1]
## [1,] -1.421085e-14
## [2,] -1.421085e-14
## [3,]  1.421085e-14
```

Theorem 11.1. *The null space of a $n \times m$ matrix \mathbf{A} is a subspace of \mathbb{R}^n .*

Proof. Do in class Show that the three requirements of the defintion of a subspace in Definition (11.1) are met.

1)

2)

3)

□

- Example: give \mathbf{A} and \mathbf{x} and determine if \mathbf{x} is in the null space of \mathbf{A} using \mathbf{R}

11.2 The basis of a subspace

Definition 11.4. A basis for a subspace \mathcal{H} of \mathcal{R}^n is

- 1) a linearly independent set in \mathcal{H} that
- 2) spans \mathcal{H} .

Equivlently, a baisis is a set of linearly independent vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ such that $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\} = \mathcal{H}$.

The requirement that the vectors of a basis are linearly independent while spanning a subspace \mathcal{H} means that a basis is a **minimal spanning set** for the subspace \mathcal{H} .

Example 11.5. Is a basis for a vector space unique?

Solution. A basis for a vector is not unique. Just like you can represent the number 16 as $1.6 * 10^1$ (base 10) or 2^4 (base 2), the basis for a vector space is also not unique.

Definition 11.5. The **standard basis** for \mathcal{R}^n is the set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ of length n where the vector \mathbf{e}_j is a vector that is 0 in every value except for a 1 in the j th position. For example,

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Notice that the matrix defined as having columns $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ is the identity matrix \mathbf{I} .

Example 11.6. What is the standard basis for \mathbb{R}^3 ?

Solution. The standard basis for \mathbb{R}^3 are the x-, y-, and z-axes. These are written as vectors where the x-axis is $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, the y-axis is $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and the z-axis is $\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$,

Example 11.7. Do the following set of vectors form a basis for \mathbb{R}^3 ?

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \text{ and } \mathbf{z} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

Solution. For the set of vectors $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, and $\mathbf{z} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ to form

a basis for \mathbb{R}^3 , we need to satisfy the two conditions in Definition (11.4). Both conditions of Definition (11.4) can be checked by combining the set of vectors \mathbf{x} , \mathbf{y} , and \mathbf{z} into a matrix and using RREF to determine the $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ and determine whether the set of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly independent. The matrix of the vectors is

$$\begin{pmatrix} 2 & -1 & 1 \\ 1 & 1 & 0 \\ 2 & 1 & 2 \end{pmatrix}$$

which is row-equivalent to

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Because the reduced row echelon form of the matrix has 3 pivot columns, the $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\} = \mathbb{R}^3$ which satisfies the first condition for being a basis. Because there is a pivot in every column, we know the set of vectors is linearly independent which satisfies the second condition for being a basis. Therefore, the set of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ forms a basis for \mathbb{R}^3 .

Using R, we can show this result first by creating the vectors \mathbf{x} , \mathbf{y} , and \mathbf{z} and then joining these into a matrix \mathbf{A} using `cbind()`

```
x <- c(2, 1, 2)
y <- c(-1, 1, 1)
z <- c(1, 0, 2)

A <- cbind(x, y, z)
```

Next, we convert \mathbf{A} to reduced row echelon form to get the row equivalent matrix

```
rref(A)
```

```
##      x y z
## [1,] 1 0 0
## [2,] 0 1 0
## [3,] 0 0 1
```

Because this reduced row echelon form has a pivot in each column we know the columns of the matrix A are linearly independent. Because there are 3 pivot columns in total, we know the span of the columns of A is \mathcal{R}^3 . Therefore, the vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ forms a basis for \mathcal{R}^3 .

Example 11.8. Do the following set of vectors form a basis for \mathcal{R}^3 ?

$$\mathbf{w} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \text{ and } \mathbf{z} = \begin{pmatrix} 4 \\ 2 \\ -2 \end{pmatrix}$$

Solution. For the set of vectors $\mathbf{w} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$, and $\mathbf{z} = \begin{pmatrix} 4 \\ 2 \\ -2 \end{pmatrix}$

to form a basis for \mathcal{R}^3 , we need to satisfy the two conditions in Definition (11.4). Both conditions of Definition (11.4) can be checked by combining the set of vectors \mathbf{w} , \mathbf{x} , \mathbf{y} , and \mathbf{z} into a matrix and using RREF to determine the $\text{span}\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ and determine whether the set of vectors $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly independent. The matrix of the vectors is

$$\begin{pmatrix} 2 & -1 & 1 & 4 \\ 1 & 1 & 0 & 2 \\ 2 & 1 & 2 & -2 \end{pmatrix}$$

which is row-equivalent to

$$\begin{pmatrix} 1 & 0 & 0 & 16/5 \\ 0 & 1 & 0 & -6/5 \\ 0 & 0 & 1 & -18/5 \end{pmatrix}$$

Because the reduced row echelon form of the matrix has 3 pivot columns, the $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\} = \mathcal{R}^3$ which satisfies the first condition for being a basis. However, there is not a pivot in every column which tells us that the set of vectors is linearly dependent which does not satisfy the second condition for being a basis. Therefore, the set of vectors $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ do not form a basis for \mathcal{R}^3 .

Using R, we can show this result first by creating the vectors `w`, `x`, `y`, and `z` and then joining these into a matrix using `cbind()`:

```
w <- c(2, 1, 2)
x <- c(-1, 1, 1)
y <- c(1, 0, 2)
z <- c(4, 2, -2)

A <- cbind(w, x, y, z)
```

Next, we convert \mathbf{A} to reduced row echelon form to get the row equivalent matrix

```
rref(A)
```

```
##      w x y   z
## [1,] 1 0 0 3.2
## [2,] 0 1 0 -1.2
## [3,] 0 0 1 -3.6
```

Because this reduced row echelon form does not have a pivot in each column, we know the columns of the matrix \mathbf{A} are linearly dependent. Because there are 3 pivot columns in total, we know the span of the columns of \mathbf{A} is \mathbb{R}^3 . Because the set of vectors $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ are linearly dependent, they do not form a basis for \mathbb{R}^3 .

Example 11.9. Do the following set of vectors form a basis for \mathbb{R}^3 ?

$$\mathbf{x} = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 3 \\ -3 \\ 4 \end{pmatrix}, \text{ and } \mathbf{z} = \begin{pmatrix} 5 \\ 9 \\ 0 \end{pmatrix}$$

Solution. For the set of vectors $\mathbf{x} = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} 3 \\ -3 \\ 4 \end{pmatrix}$, and $\mathbf{z} = \begin{pmatrix} 5 \\ 9 \\ 0 \end{pmatrix}$ to form

a basis for \mathbb{R}^3 , we need to satisfy the two conditions in Definition (11.4). Both conditions of Definition (11.4) can be checked by combining the set of vectors \mathbf{x} , \mathbf{y} , and \mathbf{z} into a matrix and using RREF to determine the $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ and determine whether the set of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly independent. The matrix of the vectors is

$$\begin{pmatrix} 4 & 3 & 5 \\ 3 & -3 & 9 \\ 2 & 4 & 0 \end{pmatrix}$$

which is row-equivalent to

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

Because the reduced row echelon form of the matrix has 2 pivot columns, the $\text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\} = \mathbb{R}^2$ which does not satisfy the first condition for being a basis

for \mathcal{R}^3 . In addition, there is not a pivot in every column which tells us that the set of vectors is linearly dependent which does not satisfy the second condition for being a basis. Therefore, the set of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ do not form a basis for \mathcal{R}^3 .

Using R, we can show this result first by creating the vectors \mathbf{x}, \mathbf{y} , and \mathbf{z} and then joining these into a matrix A using cbind():

```
x <- c(4, 3, 2)
y <- c(3, -3, 4)
z <- c(5, 9, 0)

A <- cbind(x, y, z)
```

Next, we convert A to reduced row echelon form to get the row equivalent matrix

```
rref(A)
```

```
##      x  y  z
## [1,] 1  0  2
## [2,] 0  1 -1
## [3,] 0  0  0
```

Because this reduced row echelon form does not have a pivot in each column, we know the columns of the matrix A are linearly dependent. Because there are 2 pivot columns in total, we know the span of the columns of A is \mathcal{R}^2 which is not \mathcal{R}^3 . Because the set of vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ are linearly dependent and do not span \mathcal{R}^3 , they do not form a basis for \mathcal{R}^3 .

Example 11.10. Using R, find a basis for the null space of the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 1 & 3 \\ -1 & -2 & 6 & 5 \\ 1 & 2 & -3 & 2 \end{pmatrix}$$

Solution. Given the matrix A, we look for non-trivial solutions to $\mathbf{Ax} = \mathbf{0}$

```
A <- matrix(c(2, -1, 1, 4, -2, 2, 1, 6, -3, 3, 5, 2), 3, 4)
rref(cbind(A, 0))
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    0    0    0
## [2,]    0    0    1    0    0
## [3,]    0    0    0    1    0
```

which has solution $x_1 = -2x_2$, x_2 = free, $x_3 = 0$ and $x_4 = 0$. This can be represented as a vector $\mathbf{v} = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. Thus, $\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ is a basis for the null space of \mathbf{A} .

We can check this by showing that $\mathbf{Av} = \mathbf{0}$

```
v <- c(-2, 1, 0, 0)
A %*% v
```

```
##      [,1]
## [1,]     0
## [2,]     0
## [3,]     0
```

In addition, any linear combination of the basis is also in the null space

```
A %*% (5*v)
```

```
##      [,1]
## [1,]     0
## [2,]     0
## [3,]     0
```

Theorem 11.2. *The pivot columns of a matrix \mathbf{A} form a basis for the column space of \mathbf{A} .*

Note: Use the columns of \mathbf{A} , not the columns of the matrix in echelon form.

Proof. We will provide just a sketch of the proof here. First, the column space (Definition 11.2) is the space defined by the linear combination of the columns of \mathbf{A} . Thus, the span of any subset of the vectors that make up the columns of \mathbf{A} must, by definition, be in the column space of \mathbf{A} , because any linear combination of the subset of vectors is just a linear combination of the full set of vectors with the coefficients of the vectors in the subset with the same coefficients and the coefficients of the vectors not in the subset equal to 0. Thus, the span of the subset of the vectors is contained within the span of the columns of \mathbf{A} , which is defined as the column space.

Now, choose the subset of columns of \mathbf{A} that correspond to the pivot columns of the reduced row echelon form of \mathbf{A} . If the span of the columns of $\mathbf{A} = \mathcal{R}^p$, then 1) there must be p pivot columns in the reduced row echelon form of \mathbf{A} and 2) there are p vectors in the subset of vectors where the columns of \mathbf{A}

are pivot columns. Thus, the subset of vectors spans \mathcal{R}^p (the columnspace) and, by definition, there are p pivot columns (therefore there is a pivot in each column) so the subset of vectors defined by the pivot columns of \mathbf{A} are linearly independent. \square

Example 11.11. Find a basis for the column space of the matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 2 & -3 \\ 4 & 1 & -3 & -2 \\ 4 & -1 & -3 & 1 \end{pmatrix}$$

Solution. Given the matrix \mathbf{A} , we first find its reduced row echelon form

$$\begin{pmatrix} 1 & 0 & 0 & -11/34 \\ 0 & 1 & 0 & -3/2 \\ 0 & 0 & 1 & -9/34 \end{pmatrix}$$

which has pivots in the first three columns. Thus, the column space of \mathbf{A} , which is defined as the linear combination of the vectors of \mathbf{A} , spans \mathcal{R}^3 because there are three pivot columns. The first three columns of \mathbf{A} are $\begin{pmatrix} 3 \\ 4 \\ 4 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, and $\begin{pmatrix} 2 \\ -3 \\ -3 \end{pmatrix}$. Because these 3 vectors are linearly independent (they are each pivot columns in the reduced row echelon form of \mathbf{A}) and they span \mathcal{R}^3 , they form a basis for the column space of \mathbf{A} .

Thus, any vector in the columns space of \mathbf{A} (for example, the fourth column of \mathbf{A}) can be written as a linear combination of the basis vectors of the columns space of \mathbf{A} .

Example 11.12. Find a basis for the column space of the matrix

$$\begin{pmatrix} -4 & 1 \\ 8 & -2 \\ 6 & 3 \\ 9 & 7 \end{pmatrix}$$

Solution. Given the matrix \mathbf{A} , we first find its reduced row echelon form

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

which has pivots in the first two columns. Thus, the column space of \mathbf{A} , which is defined as the linear combination of the vectors of \mathbf{A} , spans \mathcal{R}^2 because there

are two pivot columns. The first two columns of \mathbf{A} are $\begin{pmatrix} -4 \\ 8 \\ 6 \\ 9 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -2 \\ 3 \\ 7 \end{pmatrix}$.

Because these 2 vectors are linearly independent (they are each pivot columns) and they span \mathbb{R}^2 , they form a basis for the column space of \mathbf{A} .

Note however, that the first two columns of the reduced row echelon form of \mathbf{A}

(the vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$) do not form a basis for the column space of \mathbf{A} .

This can be seen because The first two columns of \mathbf{A} have non-zero entries in the 3rd and 4th elements.

Chapter 12

Dimension and Rank

```
library(dasc2594)
```

12.1 Coordinate systems

Recall the idea of polynomials (e.g., a polynomial of order p is $a_1x^p + a_2x^{p-1} + \dots + a_p x^1 + a_{p+1}x^0$) where the polynomials $x^p, x^{p-1}, \dots, x^1, x^0$ form a set of powers up to the power p of x from which the coefficients a_p, \dots, a_{p+1} can be used to make any polynomial of order p . It can be said that the powers of x ($x^p, x^{p-1}, \dots, x^1, x^0$) form a basis for all polynomials of order p .

In the previous section, we extended this analogy to vector spaces using the concept of a minimal spanning set. Consider the basis $\mathbf{b}_1, \dots, \mathbf{b}_k$ for a subspace \mathcal{H} of \mathbb{R}^n where $\text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_k\} = \mathcal{H}$. Because the set $\mathbf{b}_1, \dots, \mathbf{b}_k$ is a basis, the set of vectors is linearly independent. Then, because the set $\mathbf{b}_1, \dots, \mathbf{b}_k$ is a basis, we have the following result.

Theorem 12.1. *For each vector \mathbf{a} in the subspace \mathcal{H} of \mathbb{R}^n , and a basis $\mathbf{b}_1, \dots, \mathbf{b}_k$, there is a unique set of coefficients x_1, \dots, x_k such that*

$$\mathbf{a} = x_1\mathbf{b}_1 + \dots + x_k\mathbf{b}_k$$

Proof. In class: assume contradiction that there are two ways x_1, \dots, x_k and y_1, \dots, y_k . Show that this violates the assumption of linear dependence. \square

Definition 12.1. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ be a basis for a subspace \mathcal{H} of \mathbb{R}^n . Then, for each $\mathbf{a} \in \mathcal{H}$, the **coordinates** of \mathbf{a} with respect to the basis \mathcal{B} are the set of coefficients $\{x_1, \dots, x_k\}$ where

$$\mathbf{a} = x_1\mathbf{b}_1 + \dots + x_k\mathbf{b}_k.$$

Example 12.1. Let $\mathcal{B} = \left\{ \mathbf{b}_1 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}, \mathbf{b}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ and $\mathbf{a} = \begin{pmatrix} 5 \\ 6 \\ 1 \end{pmatrix}$. What are the coordinates of \mathbf{a} with respect to the basis \mathcal{B} ?

Solution. It can be seen that $\mathbf{a} = 3\mathbf{b}_1 - 2\mathbf{b}_2 + 0\mathbf{b}_3$ because $3 \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 1 \end{pmatrix}$. Thus the coordinates of \mathbf{a} with respect to \mathcal{B} are $\mathbf{x} = \begin{pmatrix} 3 \\ -2 \\ 0 \end{pmatrix}$

Now, the question is how to find such a solution in general. What we know is that if we write the matrix $\mathbf{B} = (\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3)$, then the coefficients for the vector \mathbf{x} are the solutions to the matrix equation

$$\mathbf{B}\mathbf{x} = \mathbf{a}$$

Notice that this is the same matrix equation as $\mathbf{Ax} = \mathbf{b}$ but written in different notation that denotes that \mathbf{B} is a basis. Because \mathbf{B} is a basis, we know that there is a pivot in every column which tells us that as long as \mathbf{a} is in the columnspace of \mathbf{B} , there will be a unique solution for the coordinates \mathbf{x} . Using an augmented matrix approach, you can solve for \mathbf{x} using elementary row operations applied to the matrix

$$\begin{array}{c} (\mathbf{B} \quad \mathbf{a}) = \begin{pmatrix} 3 & 2 & 0 & 5 \\ 0 & -3 & 0 & 6 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\ \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

which gives the solution that $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 0 \end{pmatrix}$

In R, this can be done as

```
b1 <- c(3, 0, 1)
b2 <- c(2, -3, 1)
b3 <- c(0, 0, 1)
a <- c(5, 6, 1)

rref(cbind(b1, b2, b3, a))
```

```
##      b1 b2 b3  a
```

```
## [1,] 1 0 0 3
## [2,] 0 1 0 -2
## [3,] 0 0 1 0
```

12.2 Dimension of a subspace

Definition 12.2. The **dimension** $\dim(\mathcal{H})$ of a nonzero subspace \mathcal{H} of \mathbb{R}^n is the number of (nonzero) vectors that make up a basis \mathcal{B} for \mathcal{H} . The dimension of the subspace $\mathcal{H} = \{\mathbf{0}\}$ that contains only the $\mathbf{0}$ vectors is defined as 0.

Example 12.2. Note that under this definition, the basis \mathcal{B} is not unique. For example, the following bases for the 3-dimensional subspace \mathcal{H} of \mathbb{R}^3 both have three linearly independent vectors.

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Let $\mathbf{x} = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$. What are the coordinates of \mathbf{x} with respect to \mathcal{B}_1 and \mathcal{B}_2 ?

Solution. Under the basis \mathcal{B}_1 , the coordinates of \mathbf{x} with respect to the basis \mathcal{B}_1 are $a_1 = 3$, $a_2 = 4$, and $a_3 = 0$ because

$$\mathbf{x} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + a_3 \mathbf{b}_3 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 4 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

which we write as $[\mathbf{x}]_{\mathcal{B}_1} = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$ to denote that these are the coordinates of the vector \mathbf{x} with respect to the basis \mathcal{B}_1 .

The coordinates of \mathbf{x} with respect to the basis \mathcal{B}_2 are $a_1 = 3$, $a_2 = 1$, and $a_3 = 0$ because

$$\mathbf{x} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + a_3 \mathbf{b}_3 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

which we write as $[\mathbf{x}]_{\mathcal{B}_2} = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$ to denote that these are the coordinates of the vector \mathbf{x} with respect to the basis \mathcal{B}_2 .

We can get these coordinates using R by creating augmented matrices and using row operations. For example, the coordinates of \mathbf{x} with respect to \mathcal{B}_1 are

```
B1 <- matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 1), 3, 3)
B1
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    1    0
## [3,]    0    0    1
```

```
x <- c(3, 4, 0)
x
```

```
## [1] 3 4 0
```

```
# augmented matrix
cbind(B1, x)
```

```
##           x
## [1,] 1 0 0 3
## [2,] 0 1 0 4
## [3,] 0 0 1 0
```

```
# rref of augmented matrix
rref(cbind(B1, x))
```

```
##           x
## [1,] 1 0 0 3
## [2,] 0 1 0 4
## [3,] 0 0 1 0
```

which gives the coordinates

```
rref(cbind(B1, x))[, 4]
```

```
## [1] 3 4 0
```

The coordinates of \mathbf{x} with respect to the basis \mathcal{B}_2 are

```
B2 <- matrix(c(1, 1, 0, 0, 1, 0, 1, 0, 1), 3, 3)
B2
```

```

##      [,1] [,2] [,3]
## [1,]    1    0    1
## [2,]    1    1    0
## [3,]    0    0    1

x <- c(3, 4, 0)
x

## [1] 3 4 0

# augmented matrix
cbind(B2, x)

##           x
## [1,] 1 0 1 3
## [2,] 1 1 0 4
## [3,] 0 0 1 0

# rref of augmented matrix
rref(cbind(B2, x))

```

```

##           x
## [1,] 1 0 0 3
## [2,] 0 1 0 1
## [3,] 0 0 1 0

```

which gives the coordinates

```
rref(cbind(B2, x))[, 4]
```

```
## [1] 3 1 0
```

Example 12.3. Also note that if two subspaces \mathcal{H}_1 and \mathcal{H}_2 have the same dimension (i.e., $\dim(\mathcal{H}_1) = \dim(\mathcal{H}_2) = p$), this does not mean that these are the same subspaces. For example, Let \mathcal{H}_1 and \mathcal{H}_2 be subspaces of \mathbb{R}^3 of dimension 2 with respective bases

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Note that the subspace defined by the span of the basis vectors in \mathcal{B}_1 is a plane in the x-y axes and the subspace defined by the span of the basis vectors in \mathcal{B}_2 is a plane in the x-z axes.

Exercise 12.1. What is the dimension of a basis for \mathbb{R}^n ?

12.3 Rank

Definition 12.3. The **rank** of a matrix \mathbf{A} , denoted as $\text{rank}(\mathbf{A})$, is the dimension of the column space of \mathcal{A} .

Recall that the pivot columns of \mathbf{A} form a basis for the column space of \mathbf{A} . Hence, the number of pivot columns in the matrix \mathbf{A} is the rank of the matrix \mathbf{A} .

Example 12.4. Determine the rank of the following matrices

$$1) \quad \mathbf{A} = \begin{pmatrix} -7 & 1 & -5 & 9 \\ 5 & -6 & 4 & 8 \\ -4 & -1 & -2 & 0 \end{pmatrix}$$

$$2) \quad \mathbf{B} = \begin{pmatrix} 5 & -1 & 3 \\ -6 & 4 & -5 \\ 6 & 6 & 0 \\ -7 & -25 & 9 \\ 8 & 26 & -9 \end{pmatrix}$$

$$3) \quad \mathbf{C} = \begin{pmatrix} 3 & -5 & 1 & -8 & -1 \\ -2 & -6 & -9 & 9 & -3 \\ -4 & 5 & 4 & 8 & 2 \end{pmatrix}$$

Solution. Using Definition (12.3), the rank of \mathbf{A} is equal to the dimension of the column space of \mathbf{A} where the dimension can be found by counting the number of pivot columns.

$$1) \quad \begin{pmatrix} -7 & 1 & -5 & 9 \\ 5 & -6 & 4 & 8 \\ -4 & -1 & -2 & 0 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & 200/27 \\ 0 & 1 & 0 & -34/9 \\ 0 & 0 & 1 & -349/27 \end{pmatrix} \text{ which has 3 pivot columns. Thus, } \text{rank}(\mathbf{A}) = 3$$

$$2) \quad \begin{pmatrix} 5 & -1 & 3 \\ -6 & 4 & -5 \\ 6 & 6 & 0 \\ -7 & -25 & 9 \\ 8 & 26 & -9 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ which has 2 pivot columns.}$$

Thus, $\text{rank}(\mathbf{B}) = 2$

$$3) \quad \begin{pmatrix} 3 & -5 & 1 & -8 & -1 \\ -2 & -6 & -9 & 9 & -3 \\ -4 & 5 & 4 & 8 & 2 \end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix} 1 & 0 & 0 & -465/191 & -6/191 \\ 0 & 1 & 0 & 8/191 & 42/191 \\ 0 & 0 & 1 & -93/191 & 37/191 \end{pmatrix}$$

which has 3 pivot columns. Thus, $\text{rank}(\mathbf{C}) = 3$

Theorem 12.2 (The Rank Theorem). *If a matrix \mathbf{A} has n columns, then $\text{rank}(\mathbf{A}) + \dim(\text{null}(\mathbf{A})) = n$*

Proof. The $\text{rank}(\mathbf{A})$ is number of linearly independent columns. The dimension for the $\text{null}(\mathbf{A})$ is the number of linearly dependent columns of \mathbf{A} (non-trivial solutions to $\mathbf{Ax} = \mathbf{0}$). \square

The following theorem states that any p vectors in \mathcal{R}^p that are linearly independent must span \mathcal{R}^p .

Theorem 12.3 (The Basis Theorem). *Let \mathcal{H} be a p -dimensional subspace of \mathcal{R}^n .*

- 1) *Then any linearly independent set of p elements in \mathcal{H} is a basis for \mathcal{H} .*
- 2) *Equivalently, any set of p elements of \mathcal{H} that span \mathcal{H} is a basis for \mathcal{H}*

Proof. We consider the two statements in the theorem above.

- 1) Each of the p vectors are in \mathcal{H} and the set of vectors in \mathcal{H} are linearly independent. Thus, the span of the set of vectors is \mathcal{R}^p . We have examples where two subspaces have the same dimension but are not equal, however, because each vector is in \mathcal{H} and \mathcal{H} is a subspace, all linear combinations of the vectors are in \mathcal{H} . Thus, the set of p vectors span \mathcal{H} . Thus, the set of vectors spans the subspace and are linearly independent which satisfies the conditions of Definition (11.4).
- 2) The set of p vectors span \mathcal{H} . Because \mathcal{H} is a p -dimensional subspace of \mathcal{R}^n , each vector must be linearly independent. If the vectors were not linearly independent, the p vectors would not span a p -dimensional space. Thus, the set of p vectors span \mathcal{H} . Thus, the set of vectors spans the subspace and are linearly independent which satisfies the conditions of Definition (11.4).

\square

Theorem 12.4 (Invertible Matrix Theorem (again)). *Let \mathbf{A} be a $n \times n$ matrix. Then, in addition to the current conditions from Theorem (8.5), the following statements are equivalent to \mathbf{A} being an invertible matrix:*

- a) *The columns of \mathbf{A} form a basis for \mathcal{R}^n*
- b) $\text{col}(\mathbf{A}) = \mathcal{R}^n$
- c) $\dim(\text{col}(\mathbf{A})) = n$
- d) $\text{rank}(\mathbf{A}) = n$
- e) $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$
- f) $\dim(\text{null}(\mathbf{A})) = 0$

Chapter 13

Determinants

- 3 Blue 1 Brown – The determinant

```
library(tidyverse)
library(dasc2594)
```

Definition 13.1. The determinant $\det(\mathbf{A})$ is a function of a square $n \times n$ matrix \mathbf{A} whose output is a real number that satisfies the following properties based on elementary row operations

- 1) The determinant of the $n \times n$ identity matrix \mathbf{I} is equal to 1
- 2) If a scalar multiple of one row of \mathbf{A} is added to another row of \mathbf{A} , then the determinant $\det(\mathbf{A})$ is unchanged.
- 3) Scaling a row of \mathbf{A} by a constant c multiples the determinant by c
- 4) Swapping two rows of a matrix \mathbf{A} multiplies the determinant by -1

The determinant is the unique function mapping square matrices to the real number line that satisfies the above definition.

Example 13.1. Let $\mathbf{A} = \begin{pmatrix} 5 & 3 \\ 1 & -3 \end{pmatrix}$. Find $\det(\mathbf{A})$.

Solution. We can calculate the determinant $\det(\mathbf{A})$ using row operations to get to reduced row echelon form.

$$\begin{array}{ll}
 \begin{pmatrix} 5 & 3 \\ 1 & -3 \end{pmatrix} & \det = x \\
 \text{swap row 1 and row 2} & \begin{pmatrix} 1 & -3 \\ 5 & 3 \end{pmatrix} \quad \det = -x \\
 \text{row } 2 = -5 * \text{row 1} + \text{row 2} & \begin{pmatrix} 1 & -3 \\ 0 & 18 \end{pmatrix} \quad \det = -x \\
 \text{row 2 } \div 18 & \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \quad \det = -\frac{x}{18} \\
 \text{row 1} = \text{row 1} + 3 * \text{row 2} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \det = -\frac{x}{18}
 \end{array}$$

where the last matrix is in reduced row echelon form and is the identity matrix which has determinant $\det(\mathbf{I}) = 1$. Therefore $-\frac{x}{18} = 1$ which implies that $x = -18$ so that $\det(\mathbf{A}) = -18$ Let's check our answer in R

```
A <- matrix(c(5, 1, 3, -3), 2, 2)
det(A)
```

```
## [1] -18
```

Example 13.2. Let $\mathbf{A} = \begin{pmatrix} 2 & 0 \\ -3 & 5 \end{pmatrix}$. Find $\det(\mathbf{A})$.

Solution. We can calculate the determinant $\det(\mathbf{A})$ using row operations to get to reduced row echelon form.

$$\begin{array}{ll}
 \begin{pmatrix} 2 & 0 \\ -3 & 5 \end{pmatrix} & \det = x \\
 \text{multiply row 1 by } \frac{1}{2} & \begin{pmatrix} 1 & 0 \\ -3 & 5 \end{pmatrix} \quad \det = \frac{x}{2} \\
 \text{multiply row 2 by } \frac{-1}{3} & \begin{pmatrix} 1 & 0 \\ 1 & -\frac{5}{3} \end{pmatrix} \quad \det = -\frac{x}{6} \\
 \text{row 2} = \text{row 2} - \text{row 1} & \begin{pmatrix} 1 & 0 \\ 0 & -\frac{5}{3} \end{pmatrix} \quad \det = -\frac{x}{6} \\
 \text{multiply row 2 by } -\frac{3}{5} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \det = \frac{x}{10}
 \end{array}$$

where the last matrix is in reduced row echelon form and is the identity matrix which has determinant $\det(\mathbf{I}) = 1$. Therefore $\frac{x}{10} = 1$ which implies that $x = 10$ so that $\det(\mathbf{A}) = 10$ Let's check our answer in R

```
A <- matrix(c(2, -3, 0, 5), 2, 2)
det(A)
```

```
## [1] 10
```

13.1 Determinants of 2×2 matrices

Definition 13.2. If $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2 matrix, the determinant $\det(\mathbf{A}) = ad - bc$

Example 13.3. Let $\mathbf{A} = \begin{pmatrix} 5 & 3 \\ 1 & -3 \end{pmatrix}$. What is $\det(\mathbf{A})$?

Solution. Using Definition (13.2), the determinant of $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\det(\mathbf{A}) = ad - bc = (5 * -3) - (3 * 1) = -18$

13.2 Determinants of $n \times n$ matrices

To better understand determinants of $n \times n$ matrices, we need to define the two concepts of a matrix **minor** and **cofactor**.

Definition 13.3. For an $n \times n$ matrix \mathbf{A} ,

- The (i, j) **minor** \mathbf{A}_{-i-j} is the $(n-1) \times (n-1)$ matrix obtained by deleting the i th row and the j column from \mathbf{A}
- The (i, j) **cofactor** c_{ij} is defined using the determinant of the minor where

$$c_{ij} = (-1)^{i+j} \det \mathbf{A}_{-i-j}$$

Note: The cofactor of a scalar a (1×1 matrix) is defined as $\mathbf{C}_{ij} = (-1)^{1+1} \det(a) = a$.

Note: The leading term in the cofactor definition $(-1)^{i+j}$ defines a checkerboard pattern shown below

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ + & - & + & - & \cdots \end{pmatrix}$$

Example 13.4. Let $\mathbf{A} = \begin{pmatrix} 1 & -3 & 1 \\ 4 & 2 & -3 \\ 7 & 4 & 7 \end{pmatrix}$.

- Find the minor \mathbf{A}_{-2-3}
- Find the cofactor c_{23}

Solution. Using Definition (13.3), the minor \mathbf{A}_{-2-3} is the matrix \mathbf{A} with the second row and third column removed.

Thus, the minor $\mathbf{A}_{-2-3} = \begin{pmatrix} 1 & -3 \\ 7 & 4 \end{pmatrix}$.

In R, this can be found by first defining the matrix \mathbf{A} as

```
A <- matrix(c(1, 4, 7, -3, 2, 4, 1, -3, 7), 3, 3)
```

then finding the minor \mathbf{A}_{-2-3} as

```
A_minor_23 <- A[-2, -3]
```

The matrix cofactor $c_{23} = (-1)^{2+3} \det \mathbf{A}_{-2-3} = (-1)*((1)(4) - (-3)(7)) = -25$

In R, given the minor `A_minor_23`, the cofactor is

```
(-1)^(2+3) * det(A_minor_23)
```

```
## [1] -25
```

Note that in the cofactor definition of a $n \times n$ matrix it is assumed that you can calculate the determinant of the $n-1 \times n-1$ minor \mathbf{A}_{-i-j} . From this we see that each of the $n \times n$ cofactors of \mathbf{A} are themselves the (signed) determinants of $n-1 \times n-1$ submatrices (the matrix minors). Thus, solving for all cofactors in general requires a recursive definition where smaller and smaller submatrices are evaluated.

Theorem 13.1 (Cofactor expansion). *Let \mathbf{A} be an $n \times n$ matrix with ij th elements a_{ij} . Then*

a) *The cofactor expansion along the i th row (for any fixed row i) is*

$$\det(\mathbf{A}) = \sum_{j=1}^n a_{ij} c_{ij} = a_{i1} c_{i1} + a_{i2} c_{i2} + \cdots + a_{in} c_{in}$$

b) *The cofactor expansion along the j th column (for any fixed column j) is*

$$\det(\mathbf{A}) = \sum_{i=1}^n a_{ij} c_{ij} = a_{1j} c_{1j} + a_{2j} c_{2j} + \cdots + a_{nj} c_{nj}$$

Proof. This is quite complex. For those interested, an example is available here \square

Note: The above theorem states that there are actually $2n$ ways to calculate the determinant—one for each row and column of \mathbf{A} .

Example 13.5. Use the minor/cofactor definition to calculate the determinant of the 3×3 matrix $\mathbf{A} = \begin{pmatrix} 5 & 0 & 2 \\ 1 & 3 & 3 \\ 2 & -4 & 1 \end{pmatrix}$

Solution. The determinant of \mathbf{A} can be found by using Theorem (13.1) by expanding either down a row or a column. Because the first row contains a 0, we will use the cofactor expansion theorem. The cofactor expansion along the first row is

$$\det(\mathbf{A}) = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13}$$

where a_{ij} is the ij th element of \mathbf{A} . Thus, the cofactor expansion is

$$\det(\mathbf{A}) = 5c_{11} + 0c_{12} + 2c_{13}$$

This implies that we only need to find the cofactors c_{11} and c_{13} (but not c_{12} because it gets multiplied by 0). Thus, the cofactor expansion along the first row only requires finding 2 cofactors c_{11} and c_{13} .

The minor \mathbf{A}_{-1-1} is the matrix \mathbf{A} with the first row and first column removed and is $\mathbf{A}_{-1-1} = \begin{pmatrix} 3 & 3 \\ -4 & 1 \end{pmatrix}$. The minor \mathbf{A}_{-1-3} is the matrix \mathbf{A} with the first row and third column removed and is $\mathbf{A}_{-1-3} = \begin{pmatrix} 1 & 3 \\ 2 & -4 \end{pmatrix}$. The cofactor c_{11} is given by

$$c_{11} = (-1)^{1+1}((3)(1) - (3)(-4)) = 15$$

and the cofactor c_{13} is given by

$$c_{13} = (-1)^{1+3}((1)(-4) - (3)(2)) = -10$$

Combining these, the determinant of \mathbf{A} using the cofactor expansion is

$$\begin{aligned} \det(\mathbf{A}) &= a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} \\ &= (5)(15) + (0)(c_{12}) + (2)(-10) \\ &= 55 \end{aligned}$$

Example 13.6. Use the minor/cofactor definition to calculate the determinant of the 3×3 matrix $\mathbf{A} = \begin{pmatrix} 2 & 4 & -1 \\ -3 & 0 & 2 \\ 2 & 0 & 4 \end{pmatrix}$

Solution. The determinant of \mathbf{A} can be found by using Theorem (13.1) by expanding either down a row or a column. Because the second row contains multiple zeros, we will use the cofactor expansion theorem. The cofactor expansion along the second column is

$$\begin{aligned} \det(\mathbf{A}) &= a_{12}c_{12} + a_{22}c_{22} + a_{32}c_{32} \\ &= 4c_{12} + 0c_{22} + 0c_{32} \end{aligned}$$

This implies that we only need to find the cofactor c_{12} (but not c_{22} and c_{32} because these get multiplied by 0). Thus, the cofactor expansion along the second column only requires finding the single cofactor c_{12} .

The minor \mathbf{A}_{-1-2} is the matrix \mathbf{A} with the first row and second column removed and is $\mathbf{A}_{-1-2} = \begin{pmatrix} -3 & 2 \\ 2 & 4 \end{pmatrix}$. The cofactor c_{12} is given by

$$c_{12} = (-1)^{1+2}((-3)(4) - (2,4)(2)) = -16.$$

Thus, the determinant of \mathbf{A} using the cofactor expansion is

$$\begin{aligned} \det(\mathbf{A}) &= a_{12}c_{12} + a_{22}c_{22} + a_{32}c_{32} \\ &= (4)(16) + (0)(c_{22}) + (0)(c_{32}) \\ &= 64 \end{aligned}$$

Example 13.7. Let \mathbf{A} be a $n \times n$ matrix that has all zero entries for the j th row. Find $\det(\mathbf{A})$

Solution. Using the cofactor expansion theorem (Theorem (13.1)), expand the determinant along the j th row. Then, because all entries of the j th row of \mathbf{A} are zero, this gives

$$\begin{aligned} \det(\mathbf{A}) &= a_{j1}c_{j1} + a_{j2}c_{j2} + \cdots + a_{jn}c_{jn} \\ &= 0c_{j1} + 0c_{j2} + \cdots + 0c_{jn} \\ &= 0 \end{aligned}$$

Theorem 13.2. *The determinant of a matrix \mathbf{A} is equal to the determinant of its transpose \mathbf{A}' . In other words, $\det(\mathbf{A}) = \det(\mathbf{A}')$*

Proof. Follows directly from cofactor expansion theorem. The expansion along a given row/column of \mathbf{A} is equivalent to expansion along the corresponding column/row of \mathbf{A}' (notice the row/column for \mathbf{A} got swapped to column/row for \mathbf{A}'). \square

13.3 Properties of determinants

Theorem 13.3. *A $n \times n$ square matrix \mathbf{A} is invertible if and only if $\det(\mathbf{A}) \neq 0$*

Proof. From the invertible matrix theorem (Theorem 8.5), we know that the matrix \mathbf{A} is invertible if and only if every column of \mathbf{A} is a pivot column. Therefore, each column is linearly independent from the other columns. Based on the example above, if the rows were not linearly independent, the determinant would be equal to 0 (as row operations could create a row/column of all zeros). Thus, if the determinant is not 0, the columns of \mathbf{A} are linearly independent and the matrix is invertible. \square

Theorem 13.4. *If \mathbf{A} and \mathbf{B} are $n \times n$ matrices, \mathbf{I} is an $n \times n$ identity matrix, and c is a scalar, we have*

- a) $\det(\mathbf{I}) = 1$
- b) $\det(\mathbf{A}) = \det(\mathbf{A}')$
- c) $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ if $\det(\mathbf{A}) \neq 0$ (\mathbf{A} is invertible)
- d) $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- e) $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$

13.4 Cramer's Rule and Determinants

- 3 Blue 1 Brown – Cramer's rule While commonly used for theoretical results, Cramer's rule is not commonly used in applied linear algebra. As such, we will mention Cramer's rule but not focus on it.

Theorem 13.5 (Cramer's Rule). *Let \mathbf{A} be a $n \times n$ invertible matrix. Define $\mathbf{A}_i(\mathbf{b})$ as the matrix \mathbf{A} with the i th column replaced by the vector \mathbf{b} . For example, $\mathbf{A}_i(\mathbf{b}) = (\mathbf{a}_1 \ \dots \ \mathbf{a}_{i-1} \ \mathbf{b} \ \mathbf{a}_{i+1} \ \dots \ \mathbf{a}_n)$. Then, for any $\mathbf{b} \in \mathbb{R}^n$, the unique solution to $\mathbf{Ax} = \mathbf{b}$ has entries given by*

$$x_i = \frac{\det(\mathbf{A}_i(\mathbf{b}))}{\det(\mathbf{A})} \text{ for } i = 1, 2, \dots, n$$

Example 13.8. In Cramer's rule, why do we know

- the solution is unique for any \mathbf{b} ?
- the determinant $\det(\mathbf{A}) \neq 0$?

Chapter 14

Determinants and volumes

```
library(tidyverse)
library(dasc2594)
```

Definition 14.1. The **parallelepiped** defined by n vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathcal{R}^n$ with coefficients x_1, x_2, \dots, x_n is the subset

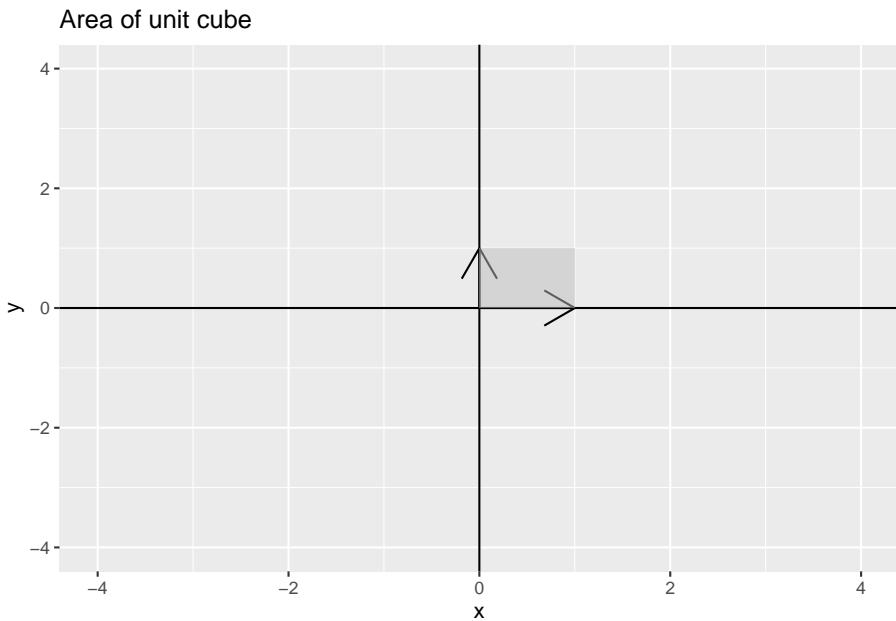
$$\mathcal{P} = \{x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n \mid 0 \leq x_1, x_2, \dots, x_n \leq 1\}$$

The determinant is a function that takes the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ that make up the columns of \mathbf{A} and returns the volume of the parallelepiped \mathcal{P} from Definition (14.1)

Example 14.1. The unit cube: in class—use standard vectors $\mathbf{e}_1, \mathbf{e}_2$, and \mathbf{e}_3

Example 14.2. parallelograms in \mathcal{R}^2 : the unit square

```
df_vector <- data.frame(x = c(1, 0), y = c(0, 1))
df_polygon <- data.frame(x = c(0, 1, 1, 0), y = c(0, 0, 1, 1))
p1 <- ggplot() +
  geom_segment(aes(x = 0, xend = df_vector$x[1], y = 0, yend = df_vector$y[1]), arrow = arrow())
  geom_segment(aes(x = 0, xend = df_vector$x[2], y = 0, yend = df_vector$y[2]), arrow = arrow())
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_polygon(data = df_polygon, aes(x = x, y = y),
               fill = "grey", alpha = 0.5) +
  ggtitle("Area of unit cube")
p1
```



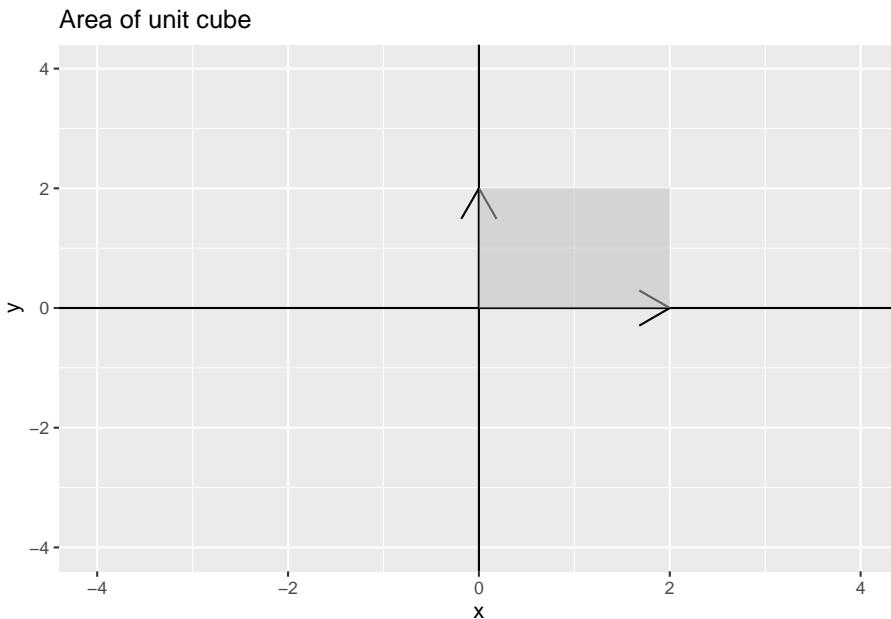
Which implies that if $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has $\det(\mathbf{A}) = 1$ because $\mathbf{A} = \mathbf{I}$ the identity matrix.

```
det(matrix(c(1, 0, 0, 1), 2, 2))
```

```
## [1] 1
```

Example 14.3. parallelograms in \mathcal{R}^2 : A larger square

```
df_vector <- data.frame(x = c(2, 0), y = c(0, 2))
df_polygon <- data.frame(x = c(0, 2, 2, 0), y = c(0, 0, 2, 2))
p1 <- ggplot() +
  geom_segment(aes(x = 0, xend = df_vector$x[1], y = 0, yend = df_vector$y[1]), arrow = arrow())
  geom_segment(aes(x = 0, xend = df_vector$x[2], y = 0, yend = df_vector$y[2]), arrow = arrow())
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  coord_cartesian(xlim = c(-4, 4), ylim = c(-4, 4)) +
  geom_polygon(data = df_polygon, aes(x = x, y = y),
               fill = "grey", alpha = 0.5) +
  ggttitle("Area of unit cube")
p1
```



Which implies that if $\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ has $\det(\mathbf{A}) = 4$ because $\mathbf{A} = 2\mathbf{I}$ and the rule is for a constant c , $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$

```
det(matrix(c(2, 0, 0, 2), 2, 2))
```

```
## [1] 4
```

The Shiny app below allows you to plot the vector for any (x, y) pair of your choosing.

```
library(shiny)
runGitHub(rep = "multivariable-math",
          username = "jtipton25",
          subdir = "shiny-apps/chapter-14/determinants-volume/")
```

Theorem 14.1 (Determinants and Volume). *Let $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ be vectors in \mathbb{R}^n , let \mathcal{P} be the parallelepiped determined by these vectors, and let \mathbf{A} be the matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Then, the absolute value of the determinant of \mathbf{A} is the volume of the parallelepiped \mathcal{P} :*

$$|\det(\mathbf{A})| = \text{volume}(\mathcal{P})$$

Proof. Recall the 4 defining properties (Definition 13.1) that characterize the determinant. These properties also characterize the absolute value of the determinant.

From the 4 defining properties, the absolute value of the determinant $|\det|$ is a function on square $n \times n$ matrices that satisfies the properties

- a) Row replacement (e.g., row $i = \text{row } i + c * \text{row } j$) of \mathbf{A} does not change $|\det(\mathbf{A})|$
- b) Scaling a row of \mathbf{A} by a scalar c changes $|\det(\mathbf{A})|$ by multiplication by $|c|$
- c) Swapping two rows of \mathbf{A} does not change $|\det(\mathbf{A})|$
- d) The determinant of the identity matrix \mathbf{I} is 1

Like the determinant and its 4 defining characteristics (Definition 13.1), the absolute value of the determinant is the only function that satisfies this relationship. Define $\text{vol}(\mathcal{P}_A)$ as the volume of the parallelepiped defined by the rows of the square matrix \mathbf{A} . In what follows, we will show that $\text{vol}(\mathcal{P}_A)$ also satisfies the 4 defining characteristics of the absolute value of the determinant and, by the uniqueness of the function of the absolute value of the determinant, is equivalent to the absolute value of the determinant.

c) We start with showing that row swaps have no impact of the volume of the parallelepiped. Swapping two rows of \mathbf{A} just reorders the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ and the order has no impact of the calculation of the volume (e.g., area in 2d = length * width = width * length)

a) Consider a row replacement of $\mathbf{a}_i \leftarrow \mathbf{a}_i + c\mathbf{a}_j$ for some $j \neq i$. Because reordering has no effect on the volume of the parallelepiped, assume WLOG (without loss of generality—fancy math speak for this one case works for all the other possible cases) that we are replacing the last row ($\mathbf{a}_n \leftarrow \mathbf{a}_n + c\mathbf{a}_j$). Then, the area of the parallelepiped is defined as the base times the height. Let the “base” be the set of vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n-1}$ which are the same for the original matrix and the row-replaced matrix. Therefore, if there is a difference in volume of the parallelepiped, it must be due to a difference in height. By definition, $\mathbf{a}_j \in \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n-1}\}$ so $c\mathbf{a}_j$ is a vector that points in the same direction as the “base” which implies that translation of \mathbf{a}_n by $c\mathbf{a}_j$ is parallel to the “base”. As this is a parallel translation, the distance from the “base” to \mathbf{a}_n must be equal to the distance from the “base” to $\mathbf{a}_n + c\mathbf{a}_j$ (the definition of parallel) which means the height is unchanged and therefore the volume of $\text{vol}(\mathcal{P}_A)$ is unchanged by row replacement.

Draw an example here

b) WLOG assume we are scaling the last row \mathbf{a}_n (we can always swap rows without changing volume so this is ok). Scaling \mathbf{a}_n by a scalar c leaves the “base” of the parallelepiped unchanged (the “base” is defined as $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n-1}$

which are unchanged). Therefore, the only question is whether the height from the base is changed when scaling \mathbf{a}_n by $|c|$. In fact, scaling the vector \mathbf{a}_n by $|c|$ changes the height by $|c|$ and therefore the volume $\text{vol}(\mathcal{P}_A)$ is scaled by $|c|$.

Draw an example in class

d) The identity matrix having volume 1 is easy. The vectors of the identity matrix \mathbf{I} define a unit cube (technically a hypercube) which has area equal to the product of the lengths of each of their sides, which are each 1.

Because the absolute value of the determinant $|\det|$ is the only function that satisfies these properties, we have

$$\text{vol}(\mathcal{P}_A) = |\det(\mathbf{A})|$$

□

Note: Because $\det(\mathbf{A}) = \det(\mathbf{A}')$, the absolute value of the determinant is equal to the volume of the parallelepiped defined by the columns of \mathbf{A} (we could just have easily done all the calculations on the columns of \mathbf{A} as the rows of \mathbf{A}).

Example 14.4. 1 by 1 matrix a (length) has $\det(a) = a$

Example 14.5. 2 by 2 matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$

Example 14.6. Find the area of a parallelogram with sides defined by the vectors $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$

area of a triangle – choose two sides and find area of parallelogram and divide by 2

Example 14.7. Recall that in data science, a probability distribution is function that has volume under the surface of one. For common distributions, particularly the normal/Gaussian distribution, the determinant is the factor that scales the function so that the volume under the surface is one.

The vector \mathbf{y} is said to have a multivariate normal distribution with mean μ and covariance matrix Σ if the probability density function of \mathbf{y} is

$$f(\mathbf{y}) = (2\pi)^{-n/2} |\det(\Sigma)|^{-1/2} e^{-\frac{1}{2}(\mathbf{y}-\mu)' \Sigma^{-1} (\mathbf{y}-\mu)}$$

Notice in this equation that the determinant $\det(\Sigma)$ plays a key role in the definition of the probability distribution. This is because

$$\int_{\mathbf{y}} (2\pi)^{-n/2} e^{-\frac{1}{2}(\mathbf{y}-\mu)' \Sigma^{-1} (\mathbf{y}-\mu)} d\mathbf{y} = |\det(\Sigma)|^{1/2}$$

which implies that

$$\frac{\int_{\mathbf{y}} (2\pi)^{-n/2} e^{-\frac{1}{2}(\mathbf{y}-\mu)' \Sigma^{-1} (\mathbf{y}-\mu)} d\mathbf{y}}{|\det(\Sigma)|^{1/2}} = 1$$

In probability and statistics, the denominator is known as the “normalizing constant.”

14.1 Volumes of Parallelpipeds

Proposition 14.1. *Let \mathbf{a}_1 and \mathbf{a}_2 be nonzero vectors. Then, for any scalar c , the area of the parallelepiped defined by \mathbf{a}_1 and \mathbf{a}_2 is the same as the area of the parallelepiped defined by the vectors \mathbf{a}_1 and $\mathbf{a}_2 + c\mathbf{a}_1$ (an elementary column operation).*

Example 14.8. Draw a parallelepiped in class. Recall that areas of a parallelepiped are defined (in 2 dimensions) as the length of the base times the height perpendicular to the base. In 3 dimensions, the volume of a parallelepiped is the base times the width (the area of the base) times the height.

14.2 Volumes of Linear Transformations

Recall linear transformations $T : \mathcal{R}^n \rightarrow \mathcal{R}^n$ (Section 6) where for any $\mathbf{x} \in \mathcal{R}^n$ (the domain), $T(\mathbf{x}) = \mathbf{A}\mathbf{x} \in \mathcal{R}^n$ (the codomain).

Theorem 14.2. *Let \mathcal{S} be a set in the domain that has a volume $\text{vol}(\mathcal{S})$. Then, the volume of the image of the set under the transformation $T(\mathcal{S})$ is $\text{vol}(T(\mathcal{S})) = |\det(\mathbf{A})| \text{vol}(\mathcal{S})$*

Chapter 15

Vector Spaces and Subspaces

- 3 Blue 1 Brown – Abstract vector spaces

Recall the definition of a subspace:

Definition 15.1. A subspace \mathcal{H} of a vector space \mathcal{V} is a subset of \mathcal{V} such that

- \mathcal{H} contains the zero vector – $\mathbf{0} \in \mathcal{H}$
- \mathcal{H} is closed under vector addition. Therefore, for \mathbf{u} and \mathbf{v} in \mathcal{H} , the sum $\mathbf{u} + \mathbf{v}$ is in \mathcal{H}
- \mathcal{H} is closed under scalar multiplication. Therefore, for \mathbf{u} in \mathcal{H} and a scalar a , the product $a\mathbf{u}$ is in \mathcal{H}

A consequence of this definition is that a subspace \mathcal{H} is closed under linear combinations.

15.1 Null space and column space

Also, recall the special subspaces of the column space and the null space.

15.1.1 Null space

Definition 15.2. The null space $\text{null}(\mathbf{A})$ of an $m \times n$ \mathbf{A} is the set of all solutions of the homogeneous equation $\mathbf{Ax} = \mathbf{0}$

Another way to write $\text{null}(\mathbf{A})$ is

$$\text{null}(\mathbf{A}) = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{Ax} = \mathbf{0}\}$$

Theorem 15.1. *The null space of an $m \times n$ matrix \mathbf{A} is a subspace of \mathbb{R}^n .*

Proof. To show that the null space of \mathbf{A} , denoted $\text{null}(\mathbf{A})$, is a subspace we need to show the following

- 1) The zero vector $\mathbf{0}$ is in the null space of \mathbf{A}
- 2) The null space of \mathbf{A} is closed under addition
- 3) The null space of \mathbf{A} is closed under scalar multiplication

The null space of \mathbf{A} is defined as the set of vectors \mathbf{x} such that $\mathbf{Ax} = \mathbf{0}$.

- 1) First, we show that the zero vector is in the subspace by setting $\mathbf{x} = \mathbf{0}$. Thus, because $\mathbf{A}\mathbf{0} = \mathbf{0}$, the zero vector $\mathbf{0}$ is in the null space of \mathbf{A} .
- 2) Next, let \mathbf{u} and \mathbf{v} be vectors in the null space of \mathbf{A} . Thus, by the definition of the null space we have $\mathbf{Au} = \mathbf{0}$ and $\mathbf{Av} = \mathbf{0}$. Consider the vector $\mathbf{u} + \mathbf{v}$ and consider $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus $\mathbf{u} + \mathbf{v}$ is in the null space of \mathbf{A} .
- 3) Finally, let \mathbf{u} be a vector in the null space of \mathbf{A} and let c be a scalar. Thus, by the definition of the null space we have $\mathbf{Au} = \mathbf{0}$. Then, consider $\mathbf{A}(c\mathbf{u}) = c\mathbf{Au} = c\mathbf{0} = \mathbf{0}$

Because the three requirements for a subspace are met, this gives us that the null space of \mathbf{A} is a subspace. \square

As a consequence, there will exist a set of vectors that span the null space $\text{null}(\mathbf{A})$. However, the null space of \mathbf{A} is defined implicitly. This means that the null space of \mathbf{A} is not obvious given the vectors of \mathbf{A} and must be checked/calculated.

Example 15.1. • In class

Find a spanning set for $\text{null}(\mathbf{A})$ where

$$\mathbf{A} = \begin{pmatrix} 7 & -2 & 7 & -4 & 5 \\ 2 & 0 & 3 & 3 & 9 \\ -5 & 2 & -5 & 7 & -2 \end{pmatrix}$$

- Find solution to system of homogeneous system of equations $\mathbf{Ax} = \mathbf{0}$

```
A <- matrix(c(7, 2, -5, -2, 0, 2, 7, 3, -5, -4, 3, 7, 5, 9, -2), 3, 5)
rref(cbind(A, 0))

##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]     1    0    0  1.50 -4.50    0
## [2,]     0    1    0  7.25  2.75    0
## [3,]     0    0    1  0.00  6.00    0
```

- Take the general solution and write as a linear combination of vectors where the coefficients are the free variables.
- general solution $x_1 = -1.5x_4 + 4.5x_5$, $x_2 = -7.25x_4 - 2.75x_5$, $x_3 = -6x_5$ and both x_4 and x_5 are free. Write out the general solution in vector form.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -1.5x_4 + 4.5x_5 \\ -7.25x_4 - 2.75x_5 \\ -6x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_4 \begin{pmatrix} -1.5 \\ -7.25 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 4.5 \\ 2.75 \\ -6 \\ 0 \\ 1 \end{pmatrix}$$

- From above, the free variables x_4 and x_5 are multiplied by the vectors $\mathbf{u} = \begin{pmatrix} -1.5 \\ -7.25 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 4.5 \\ 2.75 \\ -6 \\ 0 \\ 1 \end{pmatrix}$ where $\{\mathbf{u}, \mathbf{v}\}$ are a spanning set for the $\text{null}(\mathbf{A})$

Example 15.2. • In class – do another Find a spanning set for $\text{null}(\mathbf{A})$ where

15.1.2 Column space

Definition 15.3. The columns space $\text{col}(\mathbf{A})$ of an $m \times n$ \mathbf{A} is the set of all linear combinations of the columns of \mathbf{A} .

If $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ are the columns of \mathbf{A} , then

$$\text{col}(\mathbf{A}) = \text{span}(\mathbf{A})$$

this can be written in set notation as

$$\text{col}(\mathbf{A}) = \{\mathbf{b} : \mathbf{Ax} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathcal{R}^n\}$$

Theorem 15.2. *The column space of an $m \times n$ matrix \mathbf{A} is a subspace of \mathcal{R}^n .*

Proof. Do in class

- $\mathbf{0}$ vector
- sum of vectors
- scalar multiplication

□

Compared to the null space, the column space is defined explicitly—it is the span of the columns of \mathbf{A} . The definition of the column space results in the fact that $\text{col}(\mathbf{A})$ is the range of the linear transformation $\mathbf{x} \rightarrow \mathbf{Ax}$.

Example 15.3. • In class

Find a spanning set for $\text{col}(\mathbf{A})$ where

$$\mathbf{A} = \begin{pmatrix} 6 & 0 & 4 \\ 5 & -1 & -9 \\ -4 & 7 & 4 \\ 6 & 2 & 9 \end{pmatrix}$$

15.1.3 Understanding the differences between the column space and the null space

Contrast Between Nul A and Col A for an $m \times n$ Matrix A	
Nul A	Col A
1. Nul A is a subspace of \mathbb{R}^n .	1. Col A is a subspace of \mathbb{R}^m .
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	2. Col A is explicitly defined; that is, you are told how to build vectors in Col A .
3. It takes time to find vectors in Nul A . Row operations on $[A \quad \mathbf{0}]$ are required.	3. It is easy to find vectors in Col A . The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A .	4. There is an obvious relation between Col A and the entries in A , since each column of A is in Col A .
5. A typical vector \mathbf{v} in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	5. A typical vector \mathbf{v} in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector \mathbf{v} , it is easy to tell if \mathbf{v} is in Nul A . Just compute $A\mathbf{v}$.	6. Given a specific vector \mathbf{v} , it may take time to tell if \mathbf{v} is in Col A . Row operations on $[A \quad \mathbf{v}]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	7. Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every \mathbf{b} in \mathbb{R}^m .
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	8. Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps \mathbb{R}^n onto \mathbb{R}^m .

Chapter 16

Linearly independent sets and bases

```
library(tidyverse)
library(dasc2594)
```

Recall that a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent if the only solution to the system of equations

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$$

is the trivial solution $\mathbf{x} = \mathbf{0}$. In other words, it is not possible to write any of the vectors in the set $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ as a linear combination of the other vectors.

Definition 16.1. Let \mathcal{H} be a subspace of a vector space \mathcal{V} . Then, the set of vectors $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a **basis** for \mathcal{H} if

- The set of vectors \mathcal{B} are linearly independent
- The subspace spanned by \mathcal{B} is \mathcal{H} . In other words

$$\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \mathcal{H}$$

Example 16.1. • in class—standard basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ which are the columns of the $n \times n$ identity matrix \mathbf{I} .

Example 16.2. • in class—pick 3 vectors of length 3. Are they a basis for \mathcal{R}^3 ? What about \mathcal{R}^4 ?

Theorem 16.1 (The Spanning Set Theorem). *Let $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be vectors in the vector space \mathcal{V} and let $\mathcal{H} = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$*

- a) If one of the vectors, say \mathbf{v}_k , of \mathcal{S} is a linear combination of the remaining vectors of \mathcal{S} , then the set formed by removing the vector \mathbf{v}_k still spans \mathcal{H}
- b) If $\mathcal{H} \neq \{\mathbf{0}\}$, then some subset of \mathcal{S} spans \mathcal{H} .

Proof. a) WLOG assume \mathbf{v}_n is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ (if not, permute the labels to make the linearly dependent vector the n th vector). Then

$$\mathbf{v}_n = x_1 \mathbf{v}_1 + \dots + x_{n-1} \mathbf{v}_{n-1}$$

Because the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ span \mathcal{H} , any vector $\mathbf{b} \in \mathcal{H}$ can be written as

$$\mathbf{b} = c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n$$

for scalars c_1, \dots, c_n . Plugging the result from (1) into (1) shows that any vector \mathbf{b} in \mathcal{H} can be written only using the vectors $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$

- b) As the vectors in \mathcal{S} span \mathcal{H} , if there is a linearly dependent vector in \mathcal{S} , this vector can be removed from \mathcal{S} and the span of this subset of \mathcal{S} will still span \mathcal{H} . As long as $\mathbf{H} \neq \{\mathbf{0}\}$, there must be a least one nonzero vector in \mathbf{S} so the removing of linearly dependent vectors will stop with at least one vector. As all of the linearly dependent vectors have been removed, the subset of \mathcal{S} created in this manner will be a set of linearly independent vectors that span \mathcal{H} .

□

16.1 Bases for $\text{null}(\mathbf{A})$ and $\text{col}(\mathbf{A})$

Example 16.3. Find a basis for the $\text{col}(\mathbf{A})$ where

```
set.seed(2021)
A <- matrix(sample(-9:9, 15, replace = TRUE), 5, 3)
```

$$\mathbf{A} = \begin{pmatrix} -3 & -4 & 5 \\ -4 & -4 & -3 \\ 4 & -4 & -1 \\ -3 & 4 & 2 \\ 2 & -5 & 9 \end{pmatrix}$$

* Calculate row echelon form and identify the pivot columns. The vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ that make up the columns of \mathbf{A} that are in the pivot columns form a basis for \mathbf{A}

- Why is this? Think about the relationship between the columns of \mathbf{A} and the vector \mathbf{b} in $\mathbf{Ax} = \mathbf{b}$ that result in a consistent solution.

Example 16.4. Find a basis for the $\text{null}(\mathbf{A})$ where

```
set.seed(2021)
A <- matrix(sample(-9:9, 15, replace = TRUE), 5, 3)
```

$$\mathbf{A} = \begin{pmatrix} -3 & -4 & 5 \\ -4 & -4 & -3 \\ 4 & -4 & -1 \\ -3 & 4 & 2 \\ 2 & -5 & 9 \end{pmatrix}$$

- Calculate solutions to homogeneous system of equations, write solution in vector equation form. Vectors form a basis for $\text{null}(\mathbf{A})$
 - **note:** Facts about the basis for the null space $\text{null}(\mathbf{A})$
- 1) The spanning set produced using the method above produces a linearly independent set because the free variables are weights on the spanning vectors.
 - 2) When $\text{null}(\mathbf{A})$ contains nonzero vectors, the number of vectors in the spanning set for $\text{null}(\mathbf{A})$ is the number of free variables in the solution of $\mathbf{Ax} = \mathbf{0}$.

Example 16.5. The matrix $4 \times 5 \mathbf{A}$ has columns given by the vectors $\mathbf{a}_1, \dots, \mathbf{a}_5$ and is row equivalent to the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 3 & -2 & 1 \\ 0 & 0 & 3 & -2 & 5 \\ 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

What is a basis for $\text{col}(\mathbf{A})$ in terms of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_5$

- Note that two matrices that are row equivalent have the same linear dependence relationsihps between their vectors (but the basis for their column space is different)

Example 16.6. The matrix \mathbf{A} is row equivalent to the matrix \mathbf{B}

```
A <- matrix(c(1, 3, 2, 5, 4, 12, 8, 20, 0, 1, 1, 2, 2, 5, 3, 8, -1, 5, 2, 8), 4, 5)
B <- rref(A)
```

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{pmatrix} \mathbf{B} = \begin{pmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- What is a basis for $\text{col}(\mathbf{A})$?
- What is a basis for $\text{col}(\mathbf{B})$?
- What is $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_5)$?
- What is $\text{span}(\mathbf{b}_1, \dots, \mathbf{b}_5)$?
- Are the spaces spanned by the columns of \mathbf{A} and the columns of \mathbf{B} the same space?

Theorem 16.2. *The pivot columns of a matrix \mathbf{A} for a basis for $\text{col}(\mathbf{A})$*

Proof. sketch: \mathbf{B} rref of \mathbf{A} , linearly independent columns of \mathbf{B} are same as linearly independent columns in \mathbf{A} . Other (non-pivot) columns are linearly dependent. By spanning set theorem, non-pivot columns can be removed from the spanning set without changing the span, leaving only the pivot columns of \mathbf{A} as a basis for $\text{col}(\mathbf{A})$ \square

Chapter 17

Coordinate Systems and Dimension

- 3 Blue 1 Brown – Change of basis

```
library(tidyverse)
library(dasc2594)
```



We already know about the cartesian coordinate system (x, y, z) which has the set of basis vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

However, using the concept of a basis for a subspace \mathcal{H} of some vector space \mathcal{V} , we might want to use a different basis. Luckily, we have learned how to construct bases for $\text{col}(\mathbf{A})$ and $\text{null}(\mathbf{A})$.

You might be wondering why we want to create different bases. The usual cartesian basis has been good enough for me so far (unless you have used polar coordinates). In data science, the data often live in a high dimensional space (i.e., there are a number of data variables). However, while the data might have many variables, some of these variables are partially dependent and thus the space in which the data are embedded might be well approximated using a subspace of the original variables which can increase computation speed (less computation with fewer variables – recall from lab how the inverse of $\mathbf{X}'\mathbf{X}$ took much longer with larger numbers of variables). Thus, understanding different coordinate systems and how to change coordinate systems can lead to more efficient data representation and model fitting.

Theorem 17.1 (The unique representation Theorem). *Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for the vector space \mathcal{V} . Then, for each \mathbf{x} in \mathcal{B} , there exists a unique set of coefficients c_1, \dots, c_n such that*

$$\mathbf{x} = c_1 \mathbf{b}_1 + \dots + c_n \mathbf{b}_n$$

Proof. sketch: vectors of \mathcal{B} span \mathcal{V} so there exists a set of coefficient c_1, \dots, c_n such that

$$\mathbf{x} = c_1 \mathbf{b}_1 + \dots + c_n \mathbf{b}_n$$

is true. Assume another set of coefficients d_1, \dots, d_n exists such that

$$\mathbf{x} = d_1 \mathbf{b}_1 + \dots + d_n \mathbf{b}_n.$$

Subtract these two equations to get

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1) \mathbf{b}_1 + \dots + (c_n - d_n) \mathbf{b}_n.$$

Because \mathcal{B} is linearly independent by definition, all the weights in the equation above must be 0 (linear independence means the only solution to $\mathbf{Ax} = \mathbf{0}$ is the trivial solution). Therefore $c_i = d_i$ for $i = 1, \dots, n$. \square

Definition 17.1. Suppose $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for the vector space \mathcal{V} and $\mathbf{x} \in \mathcal{V}$. The coordinates of \mathbf{x} with respect to \mathcal{B} are the coefficients c_1, \dots, c_n such that

$$\mathbf{x} = c_1 \mathbf{b}_1 + \dots + c_n \mathbf{b}_n$$

Example 17.1. In class using standard basis in 2-dimensions and vector $\mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

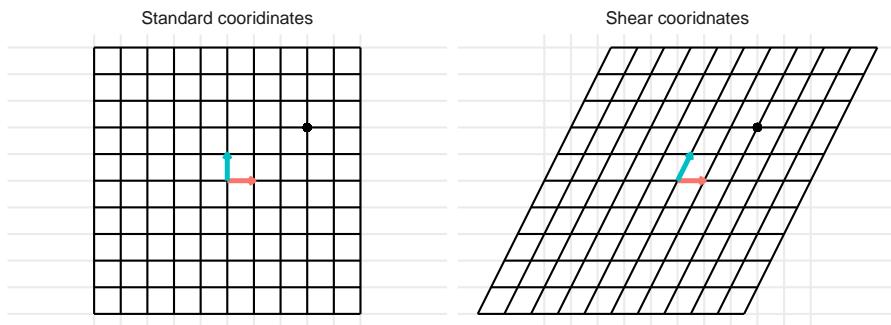
Example 17.2. In class using basis in 2-dimensions $b_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $b_2 = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$ and vector $\mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

```

transformation_matrix <- tribble(
  ~ x, ~ y,
  1, 0.5,
  0, 1) %>%
  as.matrix()

p <- plot_transformation(transformation_matrix) +
  geom_point(aes(x = 3, y = 2))
p + facet_wrap(~ time, labeller = labeller(time = c("1" = "Standard coordinates", "2"

```



17.1 Coordinates in \mathcal{R}^n

Let \mathbf{x} be defined with the standard coordinates. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis in \mathcal{R}^n . Define $\mathbf{P}_B = (\mathbf{b}_1 \ \dots \ \mathbf{b}_n)$ as the matrix with columns the vectors of the basis. Then, the coordinates $[\mathbf{x}]_B = \begin{pmatrix} [x_1]_B \\ \vdots \\ [x_n]_B \end{pmatrix}$ of \mathbf{x}

with respect to the basis \mathcal{B} can be found by solving the matrix equation

$$\mathbf{P}_B [\mathbf{x}]_B = \mathbf{x}$$

The matrix \mathbf{P}_B is called the **change-of-coordinates matrix** from \mathcal{B} to the standard basis in \mathcal{R}^n . The solution set (the coefficients) $[\mathbf{x}]_B$ can be found using row operations or by using the fact that because the columns of \mathbf{P}_B spans \mathcal{R}^n the matrix \mathbf{P}_B is invertible. Then, the coordinates of \mathbf{x} with respect to the basis \mathcal{B} is

$$[\mathbf{x}]_B = \mathbf{P}_B^{-1}\mathbf{x}$$

Theorem 17.2. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for the vector space \mathcal{V} . Then, the coordinate mapping $\mathbf{x} \rightarrow \mathbf{P}_B^{-1}\mathbf{x}$ is a one-to-one and onto transformation from \mathcal{V} to \mathcal{R}^n

Proof. First we want to show that multiplication by \mathbf{P}_B^{-1} defines a linear transformation. First, take two vectors

$$\mathbf{u} = c_1\mathbf{b}_1 + \dots + c_n\mathbf{b}_n$$

and

$$\mathbf{v} = d_1\mathbf{b}_1 + \dots + d_n\mathbf{b}_n$$

- First, we show the mapping preserves vector addition

$$\mathbf{u} + \mathbf{v} \rightarrow \mathbf{P}_B^{-1}(\mathbf{u} + \mathbf{v}) = \mathbf{P}_B^{-1}\mathbf{u} + \mathbf{P}_B^{-1}\mathbf{v}$$

which preserves vector addition

- Next, we show the mapping preserves scalar multiplication. Given scalar a ,

$$a\mathbf{u} \rightarrow \mathbf{P}_B^{-1}(a\mathbf{u}) = a\mathbf{P}_B^{-1}\mathbf{u}$$

which preserves scalar multiplication.

- Therefore, this is a linear transformation. one-to-one and onto come from fact that \mathbf{P}_B is an $n \times n$ matrix with n pivot columns (n linearly independent vectors because it is a basis for \mathcal{R}^n)

□

Example 17.3. in class—Give basis in \mathcal{R}^4 , find coefficients with respect to this basis for the vector \mathbf{x}

17.2 Dimension of a vector space

In some sense, we already know about the dimension of a vector space through the concept of a span. The span of a set of vectors defines the dimension of the vector space.

Theorem 17.3. *In a vector space \mathcal{V} with basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, any set in \mathcal{V} containing more than n vectors must be linearly dependent.*

Proof. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be a set of vectors in \mathcal{V} with $p > n$. The coordinate vectors $\{\mathbf{P}_B \mathbf{u}_1, \dots, \mathbf{P}_B \mathbf{u}_p\}$ form a linearly dependent set in \mathbb{R}^n because there are more vectors (p) than entries (n) in each vector. Thus, there exist scalars c_1, \dots, c_p , some nonzero, such that

$$c_1 \mathbf{P}_B \mathbf{u}_1 + \dots + c_p \mathbf{P}_B \mathbf{u}_p = \mathbf{0}.$$

which by linearity implies

$$\mathbf{P}_B(c_1 \mathbf{u}_1 + \dots + c_p \mathbf{u}_p) = \mathbf{0}$$

□

Because the matrix \mathbf{P}_B is a $n \times n$ matrix with n linearly independent columns, the only way the equation above can equal $\mathbf{0}$ is if the vector $c_1 \mathbf{u}_1 + \dots + c_p \mathbf{u}_p = \mathbf{0}$ (by the invertible matrix theorem). Therefore, the set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is linearly dependent because there are coefficients that allow the vectors to sum to $\mathbf{0}$. Thus, we know that for a vector space \mathcal{V} that has a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ that consists on n vectors, then every linearly independent set of vectors in \mathcal{V} contains at most n vectors.

Theorem 17.4. *If a vector space \mathcal{V} has a basis with n vectors, then every other basis of \mathcal{V} must also contain exactly n vectors.*

Proof. Let \mathcal{B}_1 be a basis of \mathcal{V} containing n vectors and let \mathcal{B}_2 be any other basis of \mathcal{V} . Because \mathcal{B}_1 and \mathcal{B}_2 are both bases, they both contain sets of linearly independent vectors. As such, the previous theorem states that each of these bases contain at most n vectors (otherwise the sets wouldn't be linearly independent). Because \mathcal{B}_2 is a basis and the basis \mathcal{B}_1 contains n vectors, \mathcal{B}_2 must contain at least n vectors. These results combined are only satisfied when \mathcal{B}_2 contains n vectors. □

Like the span defined by the columns of a matrix \mathbf{A} , there is an abstract concept called dimension which measures the “size” of a vector space.

Definition 17.2. If \mathcal{V} is spanned by a finite set of vectors, then \mathcal{V} is said to be finite dimensional. If \mathcal{V} is not spanned by a finite set of vectors, \mathcal{V} is said to be infinite dimensional. The smallest set of vectors that spans \mathcal{V} is a basis for \mathcal{V} and the number of vectors in this basis is called the **dimension** of \mathcal{V} and written as $\dim(\mathcal{V})$. If $\mathcal{V} = \{\mathbf{0}\}$, then $\dim(\mathcal{V})$ is said to be 0.

Example 17.4. in class - span of 2 or 3 linearly independent vectors

- span, dim, and geometry

Example 17.5. in class - span of 2 or 3 linearly dependent vectors

- span, dim, and geometry

17.3 Subspaces of finite dimension

Theorem 17.5. *Let \mathcal{H} be a subspace of a finite-dimensional vector space \mathcal{V} . Then, any linearly independent set in \mathcal{H} can be expanded, if necessary to form a basis for \mathcal{H} . As \mathcal{H} is a subspace of the finite-dimensional vector space \mathcal{V} , \mathcal{H} is a finite-dimensional vector space with*

$$\dim(\mathcal{H}) \leq \dim(\mathcal{V})$$

For a vector space of known dimension p , finding a basis can be simplified by finding a linearly independent set of size p .

Theorem 17.6 (The Basis Theorem). *Let \mathcal{V} be a p dimensional vector space with $p \geq 1$. Any linearly independent subset of p vectors is a basis for \mathcal{V} . Equivalently, any set of p vectors that span \mathcal{V} is automatically a basis for \mathcal{V} .*

17.4 Dimensions of $\text{null}(\mathbf{A})$ and $\text{col}(\mathbf{A})$

The dimension of $\text{null}(\mathbf{A})$ are the number of free variables in $\mathbf{Ax} = \mathbf{0}$ and the dimension of $\text{col}(\mathbf{A})$ is the number of pivot columns of \mathbf{A} .

Chapter 18

Rank

```
library(tidyverse)
library(dasc2594)
```

Definition 18.1. Given an $m \times n$ matrix \mathbf{A} , the row space, $\text{row}(\mathbf{A})$ is the set of linearly independent rows of \mathbf{A} . Thus, the row space is the span of the rows of \mathbf{A} .

Note: the row space of \mathbf{A} is the column space of the transposed matrix \mathbf{A}' .

$$\text{row}(\mathbf{A}) = \text{col}(\mathbf{A}')$$

Example 18.1. in class

Example 18.2. Find basis for row space, column space, and null space of \mathbf{A}

18.1 Rank

Definition 18.2. The rank of a matrix \mathbf{A} , $\text{rank}(\mathbf{A})$ is the dimension of the column space $\text{col}(\mathbf{A})$

Theorem 18.1 (The Rank Theorem). *Let \mathbf{A} be an $m \times n$ matrix. Then the dimension of are equal. The rank of \mathbf{A} equals the number of pivot columns of \mathbf{A} and*

$$\text{rank}(\mathbf{A}) + \dim(\text{null}(\mathbf{A})) = n$$

Proof. $\text{rank}(\mathbf{A})$ is the number of pivot columns. $\dim(\text{null}(\mathbf{A}))$ is the number of non-pivot columns. The pivot columns + non-pivot columns are the number of columns. \square

Example 18.3. in class

\mathbf{A} is an $m \times n$ matrix with $\dim(\text{null}(\mathbf{A})) = p$. What is $\text{rank}(\mathbf{A})$

Example 18.4. \mathbf{A} is a 6x9 matrix. Is it possible for $\text{null}(\mathbf{A}) = 2$?

Theorem 18.2 (Invertible Matrix Theorem + Rank). *This is an extension of the prior statement of the invertible matrix theorem 8.5 Let \mathbf{A} be an $n \times n$ matrix. Then the following statements are equivalent (i.e., they are all either simultaneously true or false).*

- 13) The columns of \mathbf{A} form a basis of \mathcal{R}^n
- 14) $\text{col}(\mathbf{A}) = \mathcal{R}^n$
- 15) $\dim(\text{col}(\mathbf{A})) = n$
- 16) $\text{rank}(\mathbf{A}) = n$
- 17) $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$
- 18) $\dim(\text{null}(\mathbf{A})) = 0$

Chapter 19

Change of basis

- 3 Blue 1 Brown – Change of basis

Consider two bases $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ for a vector space \mathcal{V} . If we have a vector $[\mathbf{x}]_B$ with coordinates in \mathcal{B} , what are the coordinates of $[\mathbf{x}]_C$ with respect to \mathcal{C} ?

- Recall: we know how to change from the standard coordinates to the basis \mathcal{B} . If \mathbf{x} is a vector in the standard coordinates and $\mathbf{P}_B = (\mathbf{b}_1 \ \dots \ \mathbf{b}_n)$ is a matrix with columns given by the basis \mathcal{B} , the coordinates of $[\mathbf{x}]_B$ of the vector \mathbf{x} with respect to the basis \mathcal{B} are

$$[\mathbf{x}]_B = \mathbf{P}_B^{-1} \mathbf{x}$$

and, as a consequence, given a vector $[\mathbf{x}]_B$ with coordinates with respect to the basis \mathcal{B} , the vector of coefficients \mathbf{x} with standard coordinates is given by

$$\mathbf{x} = \mathbf{P}_B [\mathbf{x}]_B.$$

Notice that change of coordinates is a linear transformation from \mathcal{B} to \mathcal{C} with transformation matrix \mathbf{A} . Despite the more complex notation, this is just another linear transformation [link].

Draw diagram

Now, we can combine these ideas. Given a vector $[\mathbf{x}]_B$ written with coordinates with respect to the basis \mathcal{B} , we can find the coordinates of $[\mathbf{x}]_C$ with respect to the basis \mathcal{C} . First, we find the coordinates of the vector \mathbf{x} with respect to the standard basis then find the coordinates of $[\mathbf{x}]_C$ with respect to the basis \mathcal{C} . Let $\mathbf{P}_B = (\mathbf{b}_1 \ \dots \ \mathbf{b}_n)$ and $\mathbf{P}_C = (\mathbf{c}_1 \ \dots \ \mathbf{c}_n)$, then given a vector $[\mathbf{x}]_B$ with

coordinates with respect to the basis \mathcal{B} , the coordinates $[\mathbf{x}]_C$ of this vector with respect to the basis \mathcal{C} is

$$[\mathbf{x}]_C = \mathbf{P}_C^{-1} \mathbf{P}_B [\mathbf{x}]_B.$$

Draw diagram

Example 19.1. 2-d change of basis

Example 19.2. 3-d change of basis

Chapter 20

Eigenvectors and Eigenvalues

- 3 Blue 1 Brown – Eigenvalues

```
library(tidyverse)
library(dasc2594)
set.seed(2021)
```

We have just learned about change of basis in an abstract sense. Now, we will learn about a special change of basis that is “data-driven” called an eigenvector. Eigenvectors and the corresponding eigenvalues are a vital tool in data science for data compression and modeling.

Definition 20.1. An **eigenvector** of an $n \times n$ matrix \mathbf{A} is a nonzero vector \mathbf{x} such that the matrix equation

$$\mathbf{Ax} = \lambda\mathbf{x}$$

for some scalar λ . If there exists some $\lambda \neq 0$ (a non-trivial solution), then λ is called an **eigenvalue** of \mathbf{A} corresponding to the eigenvector \mathbf{x} .

Example 20.1. It is easy to check if a vector is an eigenvalue:

Let $\mathbf{A} = \begin{pmatrix} 0 & 6 & 8 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix}$, $\mathbf{u} = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix}$, and $\mathbf{v} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$. Determine if \mathbf{u} or \mathbf{v} are eigenvectors of \mathbf{A} . If they are eigenvectors, what are the associated eigenvalues.

Now, plot \mathbf{u} , \mathbf{Au} , \mathbf{v} , and \mathbf{Av} to show this relationship geometrically.

An eigenvector is a (nonzero) vector \mathbf{x} that gets mapped to a scalar multiple of itself $\lambda\mathbf{x}$ by the matrix transformation $\mathbf{Ax} \rightarrow \lambda\mathbf{x}$. As such, we say that \mathbf{x} and

\mathbf{Ax} are collinear with the origin ($\mathbf{0}$) in the sense that these points lie on the same line that goes through the origin.

Note: The matrix \mathbf{A} must be an $n \times n$ square matrix. A similar decomposition (called the singular value decomposition) can be used for rectangular matrices.

Example 20.2. Example: reflection Draw images: <https://textbooks.math.gatech.edu/ila/eigenvalues.html>

Theorem 20.1 (The distinct eigenvalues theorem). *Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors of a matrix \mathbf{A} and suppose the corresponding eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_n$ are all distinct (different values). Then, the set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent.*

Proof. Suppose the set $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly dependent. Then, there is some j such that $\mathbf{v}_j = \sum_{k=1}^{j-1} x_k \mathbf{v}_k$. If we choose the first linearly dependent vector as j , we know that the subset of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$ is linearly independent and

$$\mathbf{v}_j = x_1 \mathbf{v}_1 + \cdots + x_{j-1} \mathbf{v}_{j-1}$$

for some scalars x_1, \dots, x_{j-1} . Multiplying the equation above on the left by \mathbf{A} on both sides gives

$$\begin{aligned} \mathbf{Av}_j &= \mathbf{A}(x_1 \mathbf{v}_1 + \cdots + x_{j-1} \mathbf{v}_{j-1}) \\ \lambda_j \mathbf{v}_j &= x_1 \mathbf{Av}_1 + \cdots + x_{j-1} \mathbf{Av}_{j-1} \\ &= x_1 \lambda_1 \mathbf{v}_1 + \cdots + x_{j-1} \lambda_{j-1} \mathbf{v}_{j-1} \end{aligned}$$

Multiplying the first equation by λ_j and subtracting this from the second equation gives

$$\mathbf{0} = \lambda_j \mathbf{v}_j - \lambda_j \mathbf{v}_j = x_1(\lambda_1 - \lambda_j) \mathbf{v}_1 + \cdots + x_{j-1}(\lambda_{j-1} - \lambda_j) \mathbf{v}_{j-1}$$

Because $\lambda_k \neq \lambda_j$ for all $k < j$, the equation above implies a linear dependence among the set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$ which is a contradiction. Therefore, our assumption that there exists a linearly dependent vector \mathbf{v}_j is violated and all the $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ are linearly independent. \square

20.1 Eigenspaces

Given a square $n \times n$ matrix \mathbf{A} , we know how to check if a given vector \mathbf{x} is an eigenvector and then how to find the eigenvalue associated with that eigenvector. Next, we want to check if a given number is an eigenvalue of \mathbf{A} and to find all the eigenvectors corresponding to that eigenvalue.

Given a square $n \times n$ matrix \mathbf{A} and a scalar λ , the eigenvectors of \mathbf{A} associated with the scalar λ (if there are eigenvectors associated with λ) are the nonzero

solutions to the equation $\mathbf{Ax} = \lambda\mathbf{x}$. This can be written as

$$\begin{aligned}\mathbf{Ax} &= \lambda\mathbf{x} \\ \mathbf{Ax} - \lambda\mathbf{x} &= \mathbf{0} \\ \mathbf{Ax} - \lambda\mathbf{Ix} &= \mathbf{0} \\ (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} &= \mathbf{0}.\end{aligned}$$

Therefore, the eigenvectors of \mathbf{A} associated with λ , if there are any, are the non-trivial solutions of the homogeneous matrix equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. In other words, the eigenvectors are the nonzero vectors in the null space $\text{null}(\mathbf{A} - \lambda\mathbf{I})$. If there is not a nontrivial solution (solution $\mathbf{x} \neq \mathbf{0}$), then λ is not an eigenvalue of \mathbf{A} .

Hey, we know how to find solutions to homogeneous systems of equations! Thus, we know how to find the eigenvectors of \mathbf{A} . All we have to do is solve the system of linear equations $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ for a given λ (actually, for all λ s, which we can't do). If only there was some way to find eigenvalues λ (hint: there is and it is coming next chapter).

Example 20.3. Let $\mathbf{A} = \begin{pmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{pmatrix}$. Then an eigenvector with eigenvalue λ is a nontrivial solution to

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

which can be written as

$$\begin{pmatrix} 3 - \lambda & 6 & -8 \\ 0 & 0 - \lambda & 6 \\ 0 & 0 & 2 - \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

which can be solved for a given λ using an augmented matrix form and row operations to reduce to reduced row echelon form.

Letting $\lambda = 3$, we have

$$\begin{pmatrix} 3 - 3 & 6 & -8 \\ 0 & 0 - 3 & 6 \\ 0 & 0 & 2 - 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

which has solution

```
lambda <- 3
rref(cbind(A - lambda * diag(nrow(A)), 0))
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     0    1    0    0
## [2,]     0    0    1    0
## [3,]     0    0    0    0
```

Definition 20.2. Let \mathbf{A} be an $n \times n$ matrix and let λ be an eigenvalue of \mathbf{A} . Then, the λ -eigenspace of \mathbf{A} is the solution set of the matrix equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ which is the subspace $\text{null}(\mathbf{A} - \lambda\mathbf{I})$.

Therefore, the λ -eigenspace is a subspace (the null space of any matrix is a subspace) that contains the zero vector $\mathbf{0}$ and all the eigenvectors of \mathbf{A} with corresponding eigenvalue λ .

Example 20.4. For $\lambda = -2, 1$, and 3 , decide if λ is a eigenvalue of the matrix $\mathbf{A} = \begin{pmatrix} 2 & -4 \\ -1 & 1 \end{pmatrix}$ and if so, compute a basis for the λ -eigenspace.

```
A <- matrix(c(2, -1, -4, 1), 2, 2)
```

- Calculate using rref on $\mathbf{A} - \lambda\mathbf{I}$ augmented with $\mathbf{0}$ and solve in parametric form
- After doing all three λ s, draw a graphic showing the λ -eigenspace basis.

20.1.1 Computing Eigenspaces

Let \mathbf{A} be a $n \times n$ matrix and let λ be a scalar.

- 1) λ is an eigenvalue of \mathbf{A} if and only if $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ has a non-trivial solution. The matrix equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ has a non-trivial solution if and only if $\text{null}(\mathbf{A} - \lambda\mathbf{I}) \neq \{\mathbf{0}\}$
- 2) Finding a basis for the λ -eigenspace of \mathbf{A} is equivalent to finding a basis for $\text{null}(\mathbf{A} - \lambda\mathbf{I})$ which can be done by finding parametric forms of the solutions of the homogeneous system of equations $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.
- 3) The dimension of the λ -eigenspace of \mathbf{A} is equal to the number of free variables in the system of equations $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ which is the number of non-pivot columns of $\mathbf{A} - \lambda\mathbf{I}$.
- 4) The eigenvectors with eigenvalue λ are the nonzero vectors in $\text{null}(\mathbf{A} - \lambda\mathbf{I})$ which are equivalent to the nontrivial solutions of $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.

Note that this leads of a fact about the 0-eigenspace.

Definition 20.3. Let \mathbf{A} be an $n \times n$ matrix. Then

- 1) The number 0 is an eigenvalue of \mathbf{A} if and only if \mathbf{A} is not invertible.
- 2) If 0 is an eigenvalue of \mathbf{A} , then the 0-eigenspace of \mathbf{A} is $\text{null}(\mathbf{A})$.

Proof. 0 is an eigenvalue of \mathbf{A} if and only if $\text{null}(\mathbf{A} - 0\mathbf{I}) = \text{null}(\mathbf{A})$. By the invertible matrix theorem, \mathbf{A} is invertible if and only if $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$ but we know that the 0-eigenspace of \mathbf{A} is not the trivial set $\{\mathbf{0}\}$ because 0 is an eigenvalue. \square

Theorem 20.2 (Invertible Matrix Theorem + eigenspaces). *This is an extension of the prior statement of the invertible matrix theorem 8.5 Let \mathbf{A} be an $n \times n$ matrix and $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the linear transformation given by $T(\mathbf{x}) = \mathbf{Ax}$. Then the following statements are equivalent (i.e., they are all either simultaneously true or false).*

- 1) \mathbf{A} is invertible.
- 2) \mathbf{A} has n pivot columns.
- 3) $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$.
- 4) The columns of \mathbf{A} are linearly independent.
- 5) The columns of \mathbf{A} span \mathbb{R}^n .
- 6) The matrix equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.
- 7) The transformation T is invertible.
- 8) The transformation T is one-to-one.
- 9) The transformation T is onto.
- 10) $\det(\mathbf{A}) \neq 0$
- 11) 0 is not an eigenvalue of \mathbf{A}

Chapter 21

The Characteristic Equation

- 3 Blue 1 Brown –

```
library(tidyverse)
library(dasc2594)
```

The characteristic equation/polynomial encodes information about the eigenvalues of the characteristic equation. In the previous chapter, we showed how we can decide if a scalar λ is an eigenvalue of a matrix and how to find the vectors associated with the eigenvalue. However, we did not learn how to find eigenvalues (other than to just randomly try λ). The characteristic equation/polynomial allows for determining the eigenvalues λ .

Definition 21.1. Let \mathbf{A} be a $n \times n$ matrix. The characteristic equation/polynomial of \mathbf{A} is the function $f(\lambda)$ given by

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

While not obvious, the function $f(\lambda)$ is a polynomial of λ but requires computing the determinant of the matrix $\mathbf{A} - \lambda\mathbf{I}$ which contains an unknown value λ .

Example 21.1. Find the characteristic equation of the matrix $\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 2 & -1 \end{pmatrix}$

- do in class

Example 21.2. Find the characteristic equation of the matrix $\mathbf{A} = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$

- do in class (expand cofactors along the third column)

Once the characteristic equation is defined, we can use the equation to solve for the eigenvalues.

Theorem 21.1. *Let \mathbf{A} be a $n \times n$ matrix and let $f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ be a characteristic polynomial. Then, the number λ_0 is an eigenvalue of \mathbf{A} if and only if $f(\lambda_0) = 0$.*

Proof. By the invertible matrix theorem, the matrix $(\mathbf{A} - \lambda_0\mathbf{I})\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if $\det(\mathbf{A} - \lambda_0\mathbf{I})\mathbf{x} = \mathbf{0}$. Therefore, the following statements are equivalent:

- λ_0 is an eigenvalue of \mathbf{A}
- $\mathbf{Ax} = \lambda_0\mathbf{x}$ has a nontrivial solution
- $(\mathbf{A} - \lambda_0\mathbf{I})\mathbf{x} = \mathbf{0}$ has a nontrivial solution
- $\mathbf{A} - \lambda_0\mathbf{I}$ is not invertible
- $\det(\mathbf{A} - \lambda_0\mathbf{I}) = \mathbf{0}$
- $f(\lambda_0) = 0$

□

Example 21.3. Using the characteristic equation of the matrix $\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 2 & -1 \end{pmatrix}$, solve for the eigenvalues and find a basis for the λ -eigenspaces

- do in class

Example 21.4. Using the characteristic equation of the matrix $\mathbf{A} = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$, solve for the eigenvalues and find a basis for the λ -eigenspaces

- do in class (expand cofactors along the third column)

21.1 Similarity

The idea behind similar matrices is to understand how the linear transformations implied by the transformation behave. Two matrices are similar if their transformation behavior (rotation, expansion/contraction, etc.) is the same but the coordinates on which the matrix operates are different.

Definition 21.2. The matrices \mathbf{A} and \mathbf{B} are said to be similar if there exists an invertible matrix \mathbf{P} where

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$$

or equivalently

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}$$

Therefore, it is possible to change \mathbf{A} into \mathbf{B} with an invertible (one-to-one and onto) transformation.

Example 21.5. Consider the following example with matrices \mathbf{A} , \mathbf{B} , and \mathbf{P} defined as below:

```
A <- matrix(c(3, 0, 0, -2), 2, 2)
A

##      [,1] [,2]
## [1,]    3    0
## [2,]    0   -2

B <- matrix(c(-12, -10, 15, 13), 2, 2)
B

##      [,1] [,2]
## [1,]   -12   15
## [2,]   -10   13

P <- matrix(c(-2, 1, 3,-1), 2, 2)
P

##      [,1] [,2]
## [1,]    -2     3
## [2,]     1    -1

P %*% B %*% solve(P)

##      [,1] [,2]
## [1,]    3    0
## [2,]    0   -2

solve(P) %*% A %*% P
```

```
##      [,1] [,2]
## [1,] -12   15
## [2,] -10   13
```

Theorem 21.2. If \mathbf{A} and \mathbf{B} are $n \times n$ similar matrices, then \mathbf{A} and \mathbf{B} will have the same characteristic polynomial and therefore the same eigenvalues.

Proof. If \mathbf{A} and \mathbf{B} are similar, then there exists an invertible matrix \mathbf{P} such that

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$$

Therefore

$$\begin{aligned}\mathbf{A} - \lambda\mathbf{I} &= \mathbf{P}\mathbf{B}\mathbf{P}^{-1} - \lambda\mathbf{I} \\ &= \mathbf{P}\mathbf{B}\mathbf{P}^{-1} - \lambda\mathbf{P}\mathbf{P}^{-1} \\ &= \mathbf{P}(\mathbf{B}\mathbf{P}^{-1} - \lambda\mathbf{P}^{-1}) \\ &= \mathbf{P}(\mathbf{B} - \lambda\mathbf{I})\mathbf{P}^{-1}\end{aligned}$$

To get the characteristic equation, we need to solve for the determinant

$$\begin{aligned}\det(\mathbf{A} - \lambda\mathbf{I}) &= \det(\mathbf{P}(\mathbf{B} - \lambda\mathbf{I})\mathbf{P}^{-1}) \\ &= \det(\mathbf{P})\det(\mathbf{B} - \lambda\mathbf{I})\det(\mathbf{P}^{-1})\end{aligned}$$

We know that $\det(\mathbf{P}^{-1}) = \frac{1}{\det(\mathbf{P})}$ (or, equivalently $\det(\mathbf{P})\det(\mathbf{P}^{-1}) = \det(\mathbf{P}\mathbf{P}^{-1}) = \det(\mathbf{I}) = 1$), we have $\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{B} - \lambda\mathbf{I})$ so that \mathbf{A} and \mathbf{B} have the same characteristic polynomial (and the same eigenvalues). \square

21.2 The geometric interpretation of similar matrices

In general, similar matrices do similar things in different spaces (different spaces in terms of different bases).

Example here

Chapter 22

Diagonalization

```
library(tidyverse)
library(dasc2594)
set.seed(2021)
```

Definition 22.1. A $n \times n$ matrix \mathbf{A} is **diagonalizable** if the matrix \mathbf{A} is similar to a diagonal matrix. This is equivalent to saying there exists some invertible $n \times n$ matrix \mathbf{P} and diagonal matrix \mathbf{D} such that

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

Example 22.1. Any diagonal matrix \mathbf{D} is diagonalizable because it is self-similar.

Theorem 22.1 (The Diagonalization Theorem). *A $n \times n$ matrix \mathbf{A} is diagonalizable if and only if the matrix \mathbf{A} has n linearly independent eigenvectors.*

In addition, the $n \times n$ matrix $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ with diagonal matrix \mathbf{D} if and only if the columns of \mathbf{P} are the linearly independent eigenvectors of \mathbf{A} . Then, the diagonal elements of \mathbf{D} are the eigenvalues of \mathbf{A} that correspond to the eigenvectors in \mathbf{P} .

Proof. This comes directly from Theorem 20.1 where if a $n \times n$ matrix has n distinct eigenvalues $\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n$, the corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are linearly independent. \square

This theorem implies that the matrix \mathbf{A} is diagonalizable if and only if the eigenvectors of \mathbf{A} form a basis for \mathcal{R}^n . When this is the case, the set of eigenvectors is called an **eigenbasis**.

Example 22.2. Consider the following example of a diagonalizable matrix \mathbf{A}

```
A <- matrix(c(9, 2, 0, -3, 2, -4, 1, 0, 3), 3, 3)
A

##      [,1] [,2] [,3]
## [1,]    9   -3    1
## [2,]    2     2    0
## [3,]    0   -4    3

eigen_A <- eigen(A)
str(eigen_A)

## List of 2
## $ values : num [1:3] 7.63 4.52 1.86
## $ vectors: num [1:3, 1:3] -0.905 -0.322 0.278 -0.407 -0.324 ...
## - attr(*, "class")= chr "eigen"

P <- eigen_A$vectors
P

##      [,1]      [,2]      [,3]
## [1,] -0.9050468 -0.4069141 -0.01938647
## [2,] -0.3217259 -0.3235720  0.27433148
## [3,]  0.2781774  0.8542377  0.96143976

D <- diag(eigen_A$values)
D

##      [,1]      [,2]      [,3]
## [1,] 7.626198 0.000000 0.000000
## [2,] 0.000000 4.515138 0.000000
## [3,] 0.000000 0.000000 1.858664

P %*% D %*% solve(P)

##      [,1]      [,2]      [,3]
## [1,] 9.000000e+00   -3    1
## [2,] 2.000000e+00     2    0
## [3,] -9.992007e-16   -4    3
```

```
all.equal(A, P %*% D %*% solve(P))
```

```
## [1] TRUE
```

Theorem 22.2. Let \mathbf{A} be a $n \times n$ diagonalizable matrix with $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. Then, the matrix power \mathbf{A}^p is

$$\mathbf{A}^p = \mathbf{P}\mathbf{D}^p\mathbf{P}^{-1}$$

Proof. In class □

Example 22.3. In this example, we apply the diagonalization theorem to the matrix \mathbf{A}

Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ which has eigenvalues 1, 2, 3. Then

the standard basis $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are corresponding eigenvectors (check the definition $\mathbf{A}\lambda = \mathbf{v}\lambda$) because

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= 1 * \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} &= 2 * \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= 3 * \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

Thus, by the diagonalization theorem, we have $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where \mathbf{P} is the identity matrix and \mathbf{D} is the diagonal matrix with entries 1, 2, 3.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

which gives us that \mathbf{A} is similar to itself.

However, there is nothing in the diagonalization theorem that says that we must put the eigenvalues in the order 1, 2, 3. If we put the eigenvalues in the order 3, 2, 1, then the corresponding eigenvectors are \mathbf{e}_3 , \mathbf{e}_2 , and \mathbf{e}_1 . Using the diagonalization theorem, we have $\mathbf{A} = \tilde{\mathbf{P}}\tilde{\mathbf{D}}\tilde{\mathbf{P}}^{-1}$ where $\tilde{\mathbf{P}}$ is the matrix with

columns \mathbf{e}_3 , \mathbf{e}_2 , and \mathbf{e}_1 and $\tilde{\mathbf{D}}$ is the diagonal matrix with entries 3, 2, 1 which results in

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^{-1}$$

which implies that the matrices $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ and $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ are similar to each other

Chapter 23

Inner product, length, and orthogonality

- 3 Blue 1 Brown – The dot product

```
library(tidyverse)
library(dasc2594)
set.seed(2021)
```

Definition 23.1. Let \mathbf{u} and \mathbf{v} be vectors in \mathcal{R}^n . Then, the **inner product** of \mathbf{u} and \mathbf{v} is $\mathbf{u}'\mathbf{v}$. The vectors \mathbf{u} and \mathbf{v} are $n \times 1$ matrices where \mathbf{u}' is a $1 \times n$ matrix and the inner product $\mathbf{u}'\mathbf{v}$ is a scalar (1×1 matrix). The inner product is also sometimes called the dot product and written as $\mathbf{u} \cdot \mathbf{v}$.

If the vectors

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

then $\mathbf{u}'\mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$

Example 23.1. Find the inner product $\mathbf{u}'\mathbf{v}$ and $\mathbf{v}'\mathbf{u}$ of

$$\mathbf{u} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 4 \\ -2 \\ 3 \end{pmatrix}$$

- do by hand

```
u <- c(2, -3, 1)
v <- c(4, -2, 3)
# u'v
sum(u*v)
```

```
## [1] 17
```

```
t(u) %*% v
```

```
##      [,1]
## [1,]    17
```

```
# v'u
sum(v*u)
```

```
## [1] 17
```

```
t(v) %*% u
```

```
##      [,1]
## [1,]    17
```

The properties of inner products are defined with the following theorem.

Theorem 23.1. Let \mathbf{u} , \mathbf{v} , and \mathbf{w} be vectors in \mathcal{R}^n and let c be a scalar. Then

- a) $\mathbf{u}'\mathbf{v} = \mathbf{v}'\mathbf{u}$
- b) $(\mathbf{u} + \mathbf{v})'\mathbf{w} = \mathbf{u}'\mathbf{w} + \mathbf{v}'\mathbf{w}$
- c) $(c\mathbf{u})'\mathbf{v} = c(\mathbf{v}'\mathbf{u})$
- d) $\mathbf{u}'\mathbf{u} \geq 0$ with $\mathbf{u}'\mathbf{u} = 0$ only when $\mathbf{u} = \mathbf{0}$

Based on the theorem above, the inner product of a vector with itself ($\mathbf{u}'\mathbf{u}$) is strictly non-negative. Thus, we can define the length of the vector \mathbf{u} (also called the **norm** of the vector \mathbf{u}).

Definition 23.2. The length of a vector $\mathbf{v} \in \mathcal{R}^n$, also called the vector **norm** $\|\mathbf{v}\|$ is defined as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

Example 23.2. Let $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathcal{R}^2$. Show that the definition of the norm satisfies the Pythagorean theorem.

Another property of the norm is how the norm changes based on scalar multiplication. Let $\mathbf{v} \in \mathcal{R}^n$ be a vector and let c be a scalar. Then $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$

Definition 23.3. A vector $\mathbf{v} \in \mathcal{R}^n$ whose length/norm is 1 is called a **unit** vector. Any vector can be made into a unit vector through **normalization** by multiplying the vector \mathbf{v} by $\frac{1}{\|\mathbf{v}\|}$ to get a unit vector $\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ in the same direction as \mathbf{v} .

23.1 Distance

In two dimensions, the Euclidean distance between the points (x_1, y_1) and (x_2, y_2) is defined as $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. In higher dimensions, a similar definition holds.

Definition 23.4. Let \mathbf{u} and \mathbf{v} be vectors in \mathcal{R}^n . Then the distance $dist(\mathbf{u}, \mathbf{v})$ between \mathbf{u} and \mathbf{v} is

$$dist(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Example 23.3. Distance between two 3-dimensional vectors

```
u <- c(3, -5, 1)
v <- c(4, 3, -2)
sqrt(sum((u-v)^2))
```

```
## [1] 8.602325
```

23.2 Orthogonal vectors

The equivalent of perpendicular lines in \mathcal{R}^n are known as orthogonal vectors.

Definition 23.5. The two vectors \mathbf{u} and \mathbf{v} in \mathcal{R}^n are orthogonal if

$$\mathbf{u}'\mathbf{v} = 0$$

23.3 Angles between vectors

Let \mathbf{u} and \mathbf{v} be vectors \mathcal{R}^n . Then, the angle between the vectors \mathbf{u} and \mathbf{v} is defined as the angle θ in the relationship

$$\mathbf{u}'\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|cos(\theta)$$

see example: `angles-as-n-gets-large.R`

23.4 Orthogonal sets

The set of vectors $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in \mathbb{R}^n is said to be an **orthogonal set** if every pair of vectors is orthogonal. In other words, for all $i \neq j$, $\mathbf{v}'_i \mathbf{v}_j = 0$. The set is called an **orthonormal set** if the set of vectors are orthogonal and for $i = 1, \dots, p$, each vector \mathbf{v}_i in the set has length $\|\mathbf{v}_i\| = 1$.

Example 23.4. Show the set of vectors $\left\{ \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \right\}$ is orthogonal

- Show these are orthogonal using R

If the set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are an orthogonal set, then the set of vectors $\left\{ \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_p}{\|\mathbf{v}_p\|} \right\}$ is an orthonormal set. Note that for each i , the length of the vector $\frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} = 1$

Theorem 23.2. Let the set $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthogonal set of nonzero vectors in \mathbb{R}^n . Then, the set of vectors in \mathcal{S} are linearly independent and therefore are a basis for the space spanned by \mathcal{S} .

Proof. Assume the set of vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ are linearly dependent. Then, there exist coefficients c_1, \dots, c_p such that

$$\mathbf{0} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p$$

Then, multiplying both equations on the left by \mathbf{v}'_1 gives

$$\begin{aligned} 0 &= \mathbf{v}'_1 \mathbf{0} = \mathbf{v}'_1(c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p) \\ &= c_1 \mathbf{v}'_1 \mathbf{v}_1 + c_2 \mathbf{v}'_1 \mathbf{v}_2 + \cdots + c_p \mathbf{v}'_1 \mathbf{v}_p \\ &= c_1 \mathbf{v}'_1 \mathbf{v}_1 + c_2 0 + \cdots + c_p 0 \\ &= c_1 \mathbf{v}'_1 \mathbf{v}_1 \end{aligned}$$

which is only equal to 0 when c_1 is equal to 0 because \mathbf{v}_1 is a nonzero vector. The above left multiplication could be repeated for each vector \mathbf{v}_i which gives all c_i must equal 0. As the only solution to the starting equation has all 0 coefficients, the set of vectors \mathcal{S} must be linearly independent. \square

A set of orthogonal vectors is called an **orthogonal basis**.

Theorem 23.3. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthogonal basis of the subspace \mathcal{W} of \mathbb{R}^n . Then for each $\mathbf{x} \in \mathcal{W}$, the coefficients for the linear combination of basis vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ for the vector \mathbf{x} are

$$\begin{aligned}\mathbf{x} &= \frac{\mathbf{x}'\mathbf{v}_1}{\mathbf{v}'_1\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}'\mathbf{v}_2}{\mathbf{v}'_2\mathbf{v}_2}\mathbf{v}_2 + \cdots + \frac{\mathbf{x}'\mathbf{v}_p}{\mathbf{v}'_p\mathbf{v}_p}\mathbf{v}_p \\ &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p\end{aligned}$$

where $c_j = \frac{\mathbf{x}'\mathbf{v}_j}{\mathbf{v}'_j\mathbf{v}_j}$. In other words, the coordinates of the vector \mathbf{x} with respect to the orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are the linear projection of the vector \mathbf{x} on the respective vectors \mathbf{v}_j .

Proof. The orthogonality of the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ gives

$$\begin{aligned}\mathbf{x}'\mathbf{v}_j &= (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p)' \mathbf{v}_j \\ &= c_j\mathbf{v}'_j\mathbf{v}_j\end{aligned}$$

Because we know that $\mathbf{v}'_j\mathbf{v}_j$ is not zero (a vector can't be orthogonal to itself), we can divide the above equality by $\mathbf{v}'_j\mathbf{v}_j$ and solve for $c_j = \frac{\mathbf{x}'\mathbf{v}_j}{\mathbf{v}'_j\mathbf{v}_j}$ \square

Thus, for a vector \mathbf{x} in the standard basis, the coordinates of \mathbf{x} with respect to an orthogonal basis can be easily calculated using dot products (rather than matrix inverses) which is an easier computation.

In fact, this is exactly the idea of using **least squares** estimation (linear regression, spline regression, etc.).

23.5 Orthogonal projections

Definition 23.6. Let \mathbf{x} be a vector in \mathcal{R}^n and let \mathcal{W} be a subspace of \mathcal{R}^n . Then the vector \mathbf{x} can be written as the **orthogonal decomposition**

$$\mathbf{x} = \mathbf{x}_{\mathcal{W}} + \mathbf{x}_{\mathcal{W}^\perp}$$

where $\mathbf{x}_{\mathcal{W}}$ is the vector in \mathcal{W} that is closest to \mathbf{x} and is called the **orthogonal projection of \mathbf{x} onto \mathcal{W}** and $\mathbf{x}_{\mathcal{W}^\perp}$ is the orthogonal projection of \mathbf{x} onto \mathcal{W}^\perp , the subspace \mathcal{W}^\perp of \mathcal{R}^n that is complementary to \mathcal{W} and is called the **orthogonal complement**.

Draw picture in class - \mathcal{W} is a plane, orthogonal projection of a vector onto the plane

This leads to the projection theorem that decomposes a vector $\mathbf{x} \in \mathcal{R}^n$ into components that are

Theorem 23.4. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthogonal basis of the subspace \mathcal{W} of \mathbb{R}^n . Then for each $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of \mathbf{x} onto \mathcal{W} is given by

$$\begin{aligned}\mathbf{x}_{\mathcal{W}} &= \frac{\mathbf{x}' \mathbf{v}_1}{\mathbf{v}'_1 \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{x}' \mathbf{v}_2}{\mathbf{v}'_2 \mathbf{v}_2} \mathbf{v}_2 + \cdots + \frac{\mathbf{x}' \mathbf{v}_p}{\mathbf{v}'_p \mathbf{v}_p} \mathbf{v}_p \\ &= c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p\end{aligned}$$

where the coefficient c_j corresponding to the vector \mathbf{v}_j of the linear combination of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is given by $c_j = \frac{\mathbf{x}' \mathbf{v}_j}{\mathbf{v}'_j \mathbf{v}_j}$. In other words, the coordinates of the vector \mathbf{x} with respect to the orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are the linear projection of the vector \mathbf{x} on the respective vectors \mathbf{v}_j .

You might be wondering what use orthogonal projections are. In fact, linear regression (and most common regression models) use orthogonal projections to fit a line (or surface) of best fit. This leads to the important theorem that allows us to project a vector $\mathbf{y} \in \mathbb{R}^n$ onto the column space of a $n \times p$ matrix \mathbf{X} (which is exactly the linear regression of \mathbf{y} onto \mathbf{X}).

Theorem 23.5 (Orthogonal Projection Theorem). Let \mathbf{X} be a $n \times p$ matrix, let $\mathcal{W} = \text{col}(\mathbf{X})$, and let \mathbf{y} be a vector in \mathbb{R}^n . Then the matrix equation

$$\mathbf{X}' \mathbf{X} \beta = \mathbf{X}' \mathbf{y}$$

with respect to the unknown coefficients β is consistent and $\mathbf{y}_{\mathcal{W}} = \mathbf{X} \beta$ for any solution β .

Proof. Show this in class □

In addition, if the columns of \mathbf{X} are linearly independent, then the coefficients β are given by

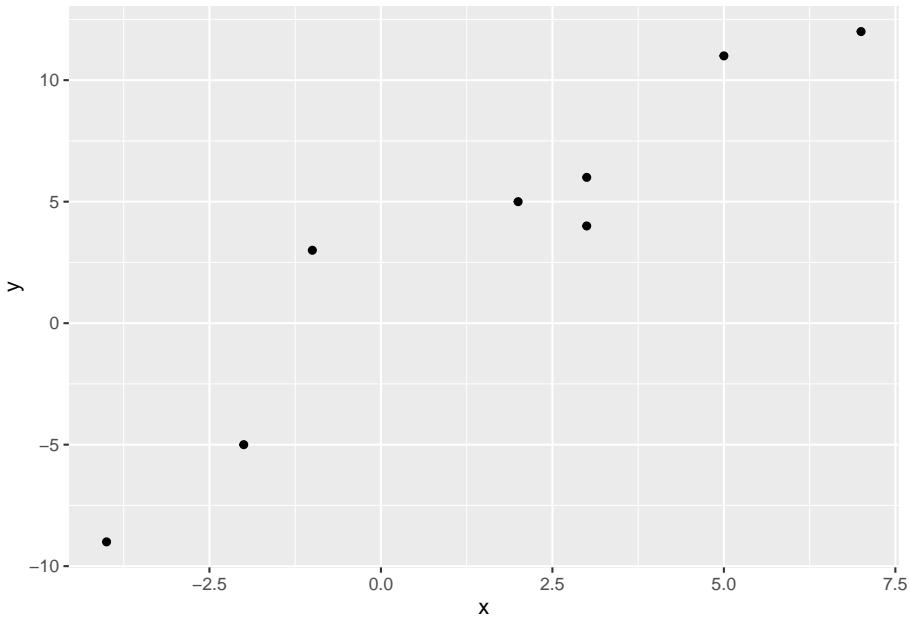
$$\beta = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

which is the least squares solution to the linear regression problem. For example, let \mathbf{X} and \mathbf{y} be defined as below

```
X <- cbind(1, c(2, -1, 3, -4, 5, 7, -2, 3))
y <- c(5, 3, 4, -9, 11, 12, -5, 6)
```

Plotting this data shows the strong positive linear relationship

```
# The first column is a basis for a constant term (the intercept)
data.frame(x = X[, 2], y = y) %>%
  ggplot(aes(x = x, y = y)) +
  geom_point()
```



We can solve for the coefficients **beta** using the linear project theorem

```
beta <- solve(t(X) %*% X) %*% t(X) %*% y
```

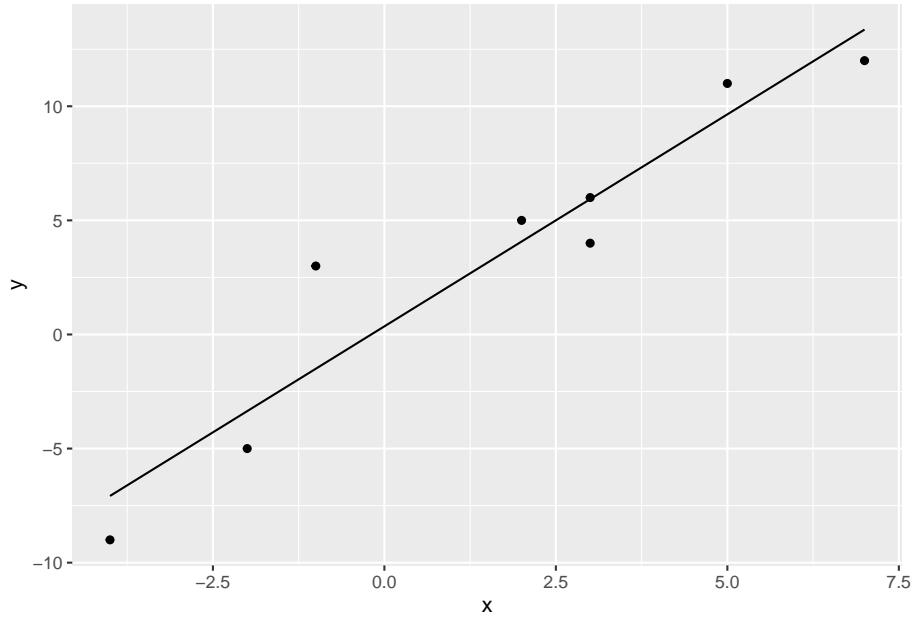
and using this solution, solve for the projection \mathbf{y}_W of \mathbf{y} onto \mathbf{X}

```
y_W <- X %*% beta
```

Plotting the projection \mathbf{y}_W gives

```
# The first column is a basis for a constant term (the intercept)
data.frame(x = X[, 2], y = y, y_W = y_W) %>%
  ggplot(aes(x = x, y = y)) +
  geom_point() +
  geom_line(aes(x = x, y = y_W))
```

202 CHAPTER 23. INNER PRODUCT, LENGTH, AND ORTHOGONALITY

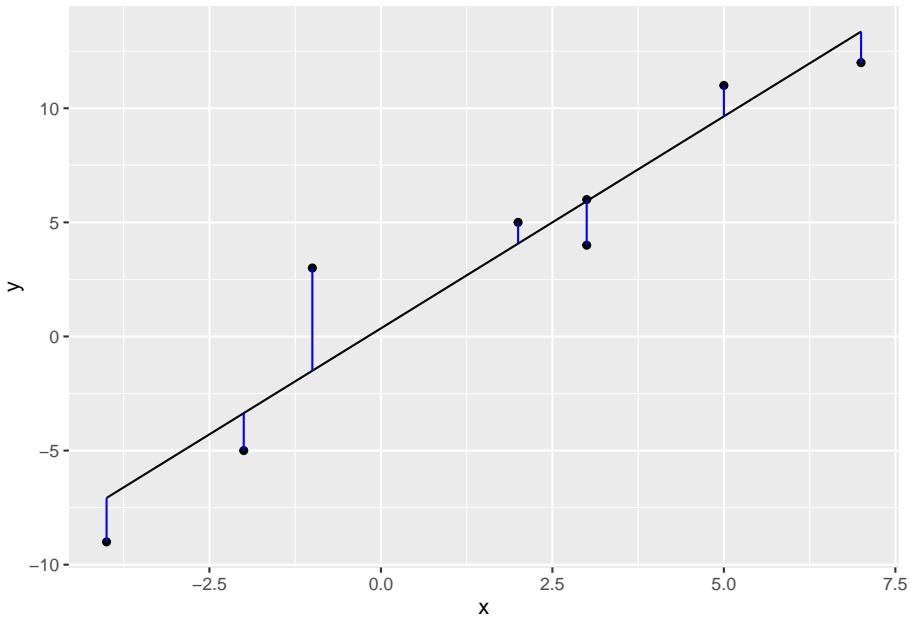


The complement of the project (called the **residuals** in statistics) is given by
 $\mathbf{y}_{\mathcal{W}^\perp} = \mathbf{y} - \mathbf{y}_{\mathcal{W}}$

```
y_W_perp <- y - y_W
```

and can be visualized as the orthogonal projection using segments

```
# The first column is a basis for a constant term (the intercept)
data.frame(x = X[, 2], y = y, y_W = y_W) %>%
  ggplot(aes(x = x, y = y)) +
  geom_point() +
  geom_line(aes(x = x, y = y_W)) +
  geom_segment(aes(x = x, y = y_W, xend = x, yend = y_W + y_W_perp), color = "blue")
```



Recall that the orthogonal projection gives the “closest” vector \mathbf{y}_W to \mathbf{y} that is in the subspace \mathcal{W} that is the span of the column space of \mathbf{X} . See <https://www.enchufa2.es/archives/least-squares-as-springs-the-shiny-app.html> for an example.

Chapter 24

Graphs and Limits

```
library(tidyverse)
library(dasc2594)
set.seed(2021)
```

Here we start a transition to topics in vector calculus. We will start with a discussion of functions of two variables (although the functions are not assumed to be linear). We define a function of two variables explicitly as $z = f(x, y)$.

Definition 24.1. Like with linear functions, we can define the domain and range for general functions of two variables. A function $f(x, y)$ assigns each point (x, y) in some **domain** \mathcal{D} in \mathbb{R}^2 to a unique number z in a subset of \mathbb{R} . The set of inputs \mathcal{D} is called the **domain** of the function and the **range** is the set of real numbers z that are the output of the function over all the inputs in \mathcal{D}

Example 24.1. Let $f(x, y) = \sqrt{1 - x^2 - y^2}$.

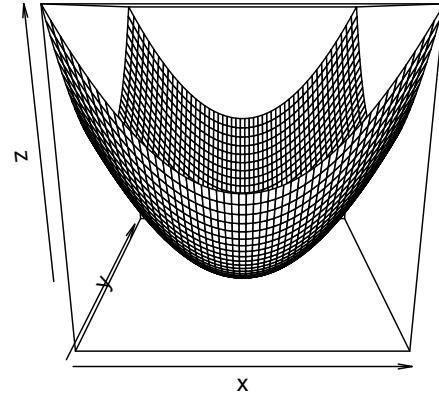
- The domain of f is the set of points (x, y) such that $x^2 + y^2 \leq 1$ which is the unit circle (**draw picture**).
- The range is the unit interval $[0, 1]$

24.1 Graphs and level curves

Example 24.2. The parabola

Consider the function of two variables

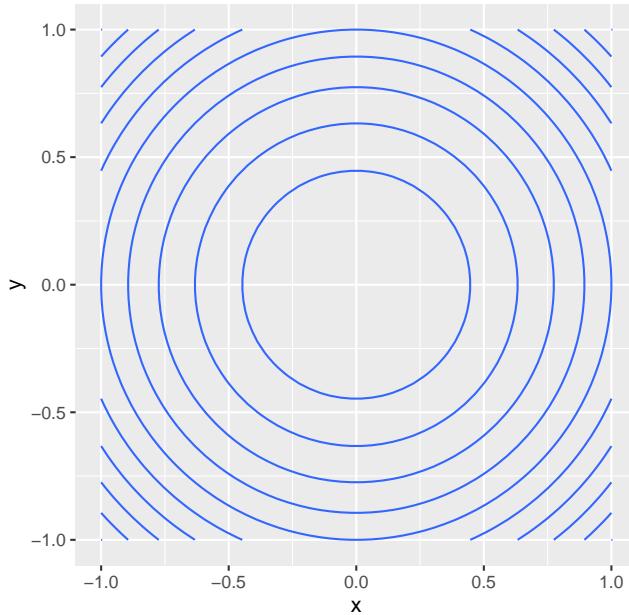
$$f(x, y) = x^2 + y^2$$



which defines the surface

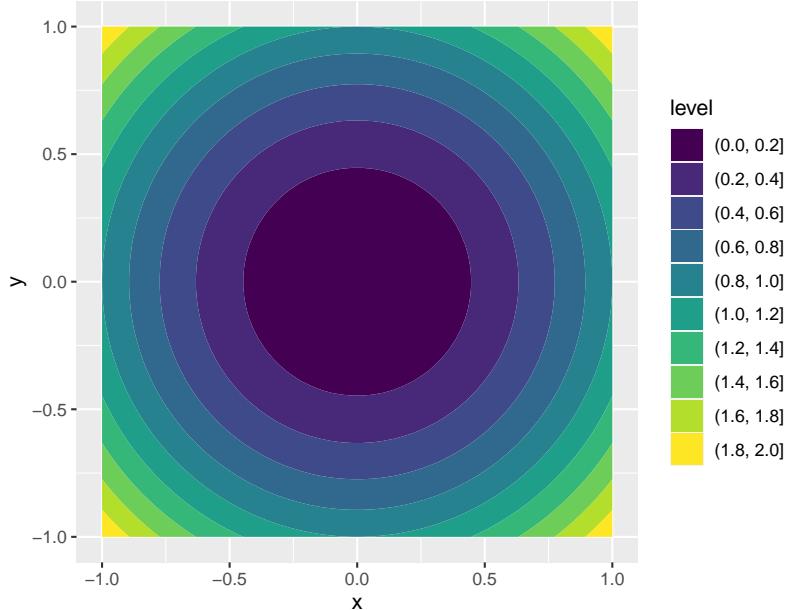
The 3-dimensional surface can be represented in 2-dimensions using level curves
(think of a topographic map)

```
data.frame(expand.grid(x, y)) %>%
  rename(x = Var1, y = Var2) %>%
  mutate(z = parabola(x, y)) %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour() +
  coord_fixed(ratio = 1)
```



where each curve in the (x, y) plane has exactly the same value of $f(x, y)$. Alternatively, this can be represented using filled level curves

```
data.frame(expand.grid(x, y)) %>%
  rename(x = Var1, y = Var2) %>%
  mutate(z = parabola(x, y)) %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour_filled() +
  coord_fixed(ratio = 1)
```

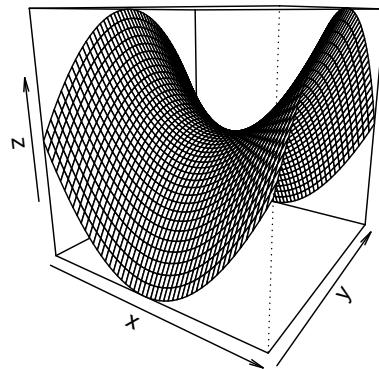


Notice that although the original parabola was continuous, these 2-d representations simplify the diagram by representing the contours as discrete values.

Example 24.3. A saddle function

Consider the function of two variables

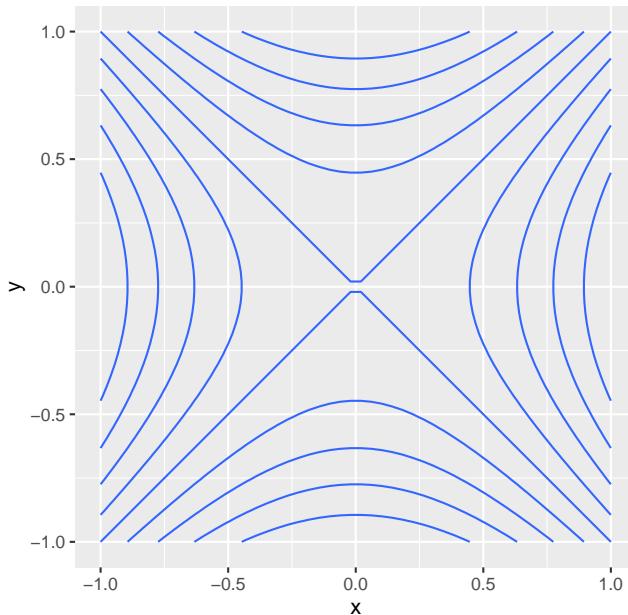
$$f(x, y) = x^2 - y^2$$



which defines the surface

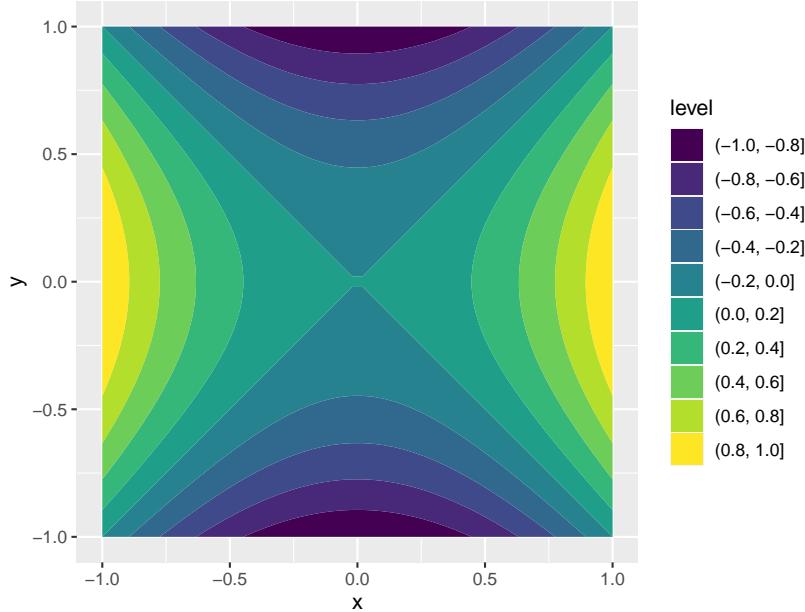
The 3-dimensional surface can be represented in 2-dimensions using level curves
(think of a topographic map)

```
data.frame(expand.grid(x, y)) %>%
  rename(x = Var1, y = Var2) %>%
  mutate(z = saddle(x, y)) %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour() +
  coord_fixed(ratio = 1)
```



where each curve in the (x, y) plane has exactly the same value of $f(x, y)$.
 Alternatively, this can be represented using filled level curves

```
data.frame(expand.grid(x, y)) %>%
  rename(x = Var1, y = Var2) %>%
  mutate(z = saddle(x, y)) %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour_filled() +
  coord_fixed(ratio = 1)
```



Notice that although the original saddle was continuous, these 2-d representations simplify the diagram by representing the contours as discrete values.

24.2 Limits

For functions of several variables, we have to define limits and continuity for these multivariable settings. For now, we focus on two variable functions as the multivariable case follows similar from the two variable case.

Let $P(x, y) \rightarrow P_0(a, b)$ be a path in the $x-y$ plane that starts at the point $P(x, y)$ and ends at the point $P_0(a, b)$ with coordinates (a, b) . Thus, we can understand the limit as the fixed value of $f(x, y)$ for which all paths that connect the points $P(x, y)$ that are “close” to $P_0(a, b)$ converge to.

For one-dimensional limits, “close” was defined as distance. Thus, for multivariable functions, “close” is defined as the Euclidean distance defined by a “ball” of radius δ and the limits examines the function output as the radius δ goes to 0.

Recall that the distance $dist((x, y), (a, b))$ between two points (x, y) and (a, b) is

$$dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$$

** Draw images**

Definition 24.2 (Limit of a function of two variables). The function $f(x, y)$ has limit L as $P(x, y) \rightarrow P_0(a, b)$, written

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = \lim_{P \rightarrow P_0} f(x, y) = L,$$

if, for any $\epsilon > 0$ (the radius of the ball that defines the “closeness” of the point), there exists a $\delta > 0$ such that

$$|f(x, y) - L| < \epsilon$$

whenever (x, y) is in the domain of f and

$$0 < \sqrt{(x - a)^2 + (y - b)^2} < \delta.$$

In the definition, as the value of δ is getting smaller, the distance between the set of points $P(x, y)$ within radius δ of the point $P_0(a, b)$ is getting smaller. As a consequence, the limit in the definition above exists only if $f(x, y)$ approaches the value L along **all possible paths** in the domain of f .

Example 24.4. In class notes * Future work: write out hand-written examples

Theorem 24.1. Let L and M be real numbers and let $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = L$ and $\lim_{(x,y) \rightarrow (a,b)} g(x, y) = M$ for functions $f(x, y)$ and $g(x, y)$. Let c be a constant and $n > 0$, then:

1) *Sum of limits:*

$$\lim_{(x,y) \rightarrow (a,b)} (f(x, y) + g(x, y)) = L + M$$

2) *Difference of limits:*

$$\lim_{(x,y) \rightarrow (a,b)} (f(x, y) - g(x, y)) = L - M$$

3) *Scalar multiple of the limit:*

$$\lim_{(x,y) \rightarrow (a,b)} cf(x, y) = cL$$

4) *Product of limits:*

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y)g(x, y) = LM$$

5) *Quotient of limits:* As long as $M > 0$ we have

$$\lim_{(x,y) \rightarrow (a,b)} \frac{f(x, y)}{g(x, y)} = \frac{L}{M}$$

6) *Power of the limit:*

$$\lim_{(x,y) \rightarrow (a,b)} f(x,y)^n = L^n$$

7) *Root of the limit:* If n is even, we assume $L > 0$

$$\lim_{(x,y) \rightarrow (a,b)} f(x,y)^{1/n} = L^{1/n}$$

Example 24.5. Use the rules above to evaluate the limit $\lim_{(x,y) \rightarrow (2,3)} 4x^3y + \sqrt{xy}$

24.2.1 Boundary points

Definition 24.3. Define a region \mathcal{D} in \mathbb{R}^2 .

- An **interior** point P of \mathcal{D} is a point that lies entirely in the region \mathcal{D} . Mathematically, a point P is an interior point of \mathcal{D} if it is possible to define a ball of radius $\epsilon > 0$ centered at P such that this ball only contains points within \mathcal{D} .
- A **boundary** point P of \mathcal{D} is a point that lies on the edge of the region \mathcal{D} . Mathematically, a point P is an boundary point of \mathcal{D} if every ball of radius $\epsilon > 0$ centered at P contains at least one point in \mathcal{D} and one point outside \mathcal{D} .

Example 24.6. Let $f(x,y) = \sqrt{1 - x^2 - y^2}$. The boundary points are all the points on the unit circle and the interior points are all the points on the interior of the unit disk. **draw picture**

Definition 24.4. A region \mathcal{D} is said to be **open** if it only contains interior points (i.e., \mathcal{D} has no boundary points). A region \mathcal{D} is said to be **closed** if the region contains all its boundary points.

Figure: limit paths along the boundary

Example 24.7. Consider the $\lim_{(x,y) \rightarrow (4,4)} \frac{x^2 - y^2}{x - y}$. Because the point $(4,4)$ is not a valid point in the domain (can't divide by $4 - 4 = 0$), the point $(4,4)$ is a boundary point of the domain. The boundary of the domain that is not contained in the domain is the set of points $x = y$. Assuming we are not taking a path along the boundary, we know that $x \neq y$ in the interior of the domain. Hence,

$$\begin{aligned} \lim_{(x,y) \rightarrow (4,4)} \frac{x^2 - y^2}{x - y} &= \lim_{(x,y) \rightarrow (4,4)} \frac{(x-y)(x+y)}{x-y} \\ &= \lim_{(x,y) \rightarrow (4,4)} x + y = 4 + 4 = 8 \end{aligned}$$

for all paths that do not cross the line $y = x$.

Example 24.8. nonexistence of limit in class

24.3 Continuity

A very important property of functions is continuity. In a general sense, a function is continuous if two nearby input values result in nearby output values. As a graph, this means that there are no hops, skips, or jumps.

Definition 24.5. The function $f(x, y)$ is said to be continuous at the point (a, b) if the following are true

- 1) $f(a, b)$ is defined at (a, b)
- 2) $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$ exists
- 3) $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$

Example 24.9. checking continuity in class $f(x, y) = \begin{cases} \frac{x^2-y^2}{x-y} & \text{if } x \neq y \\ x + y & \text{if } x = y \end{cases}$

Chapter 25

Partial Derivatives

```
library(tidyverse)
library(plotly)
library(dasc2594)
set.seed(2021)
```

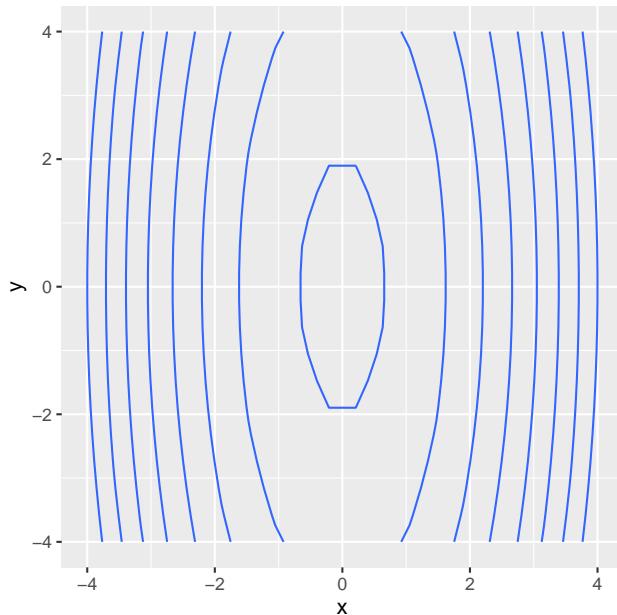
Recall that for a function of one variable, the derivative gives the rate of change of the function with respect to that variable. The function $f(x)$ has an instantaneous rate of change $\frac{d}{dx}f(x)$, assuming the derivative $\frac{d}{dx}f(x)$ exists.

This concept can be extended to functions of multivariables where we now have to specify a direction in which the function changes. For example, consider a mountain which is very steep in the north/south direction but is much less steep in the east/west direction. Thus, the directional derivative in the north/south direction will have a larger absolute value (higher rate of change) than the directional derivative in the east/west direction.

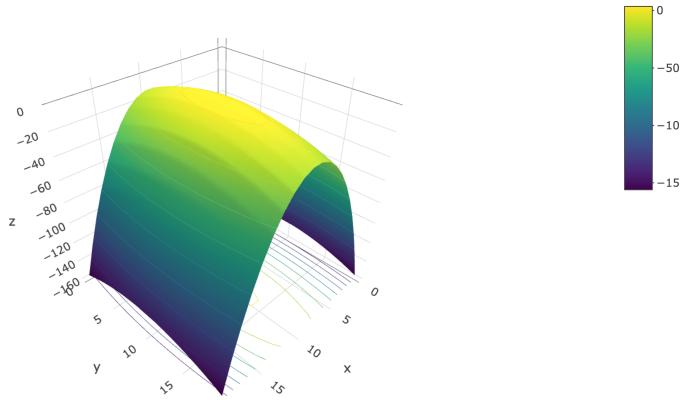
For example, consider the function

```
mountain <- function(x, y) {
  4 - 9 * x^2 - y^2
}
dat <- expand_grid(x = seq(-4, 4, length.out = 20), y = seq(-4, 4, length.out = 20)) %>%
  mutate(z = mountain(x, y))

dat %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour() +
  coord_fixed(ratio = 1)
```



```
plot_ly(z = matrix(dat$z, 20, 20)) %>%
  add_surface(
    contours = list(
      z = list(
        show=TRUE,
        usecolormap=TRUE,
        highlightcolor="#ff0000",
        project=list(z=TRUE)
      )
    )
  )
```



Note that the partial derivative asks the question “What is the rate of change of one variable holding all the other variables constant?” Thus, the question can be phrased as what is the derivative of the function $f(x, y)$ at the point (a, b) where we only let one variable change? To make the notation of a partial derivative clear, a special symbol is used where $\frac{\partial}{\partial x}$ is the partial derivative with respect to the x variable (holding the y variable constant).

draw surfaces with marginal slices

Definition 25.1. The partial derivative of the function $f(x, y)$ with respect to x at the point (a, b) is

$$\frac{\partial}{\partial x} f(x, y) = f_x(x, y) = \lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h}.$$

The partial derivative of the function $f(x, y)$ with respect to y at the point (a, b) is

$$\frac{\partial}{\partial y} f(x, y) = f_y(x, y) = \lim_{h \rightarrow 0} \frac{f(a, b + h) - f(a, b)}{h},$$

as long as these limits exist.

Example 25.1. partial derivatives using limit definition

Example 25.2. Let $f(x, y) = 3x^2 - 4y^3 + 3$, Compute $\frac{\partial}{\partial x} f(x, y)$ and $\frac{\partial}{\partial y} f(x, y)$. Then evaluate each derivative at $(-2, 3)$.

Notice that you can find partial derivatives by holding all the other variables constant and then finding the equivalent univariate derivative.

25.1 Higher-order partial derivatives

We can calculate the partial derivatives of partial derivatives. The derivatives could be with respect to the same variable repeatedly or the derivatives could

be with respect to different variables in which case we call these **mixed partial derivatives**. Notation for higher order partial derivatives is $\frac{\partial^2}{\partial x \partial y} f(x, y) = f_{xy}(x, y)$ which says first take the partial derivative of f with respect to y then take the partial derivative with respect to x . The possible sets of second-order partial derivatives for functions of two variables are shown in the table below

Notation 1	Notation 2
$\frac{\partial}{\partial x} \frac{\partial}{\partial x} f(x, y) = \frac{\partial^2}{\partial x^2} f(x, y)$	$f_{xx}(x, y)$
$\frac{\partial}{\partial y} \frac{\partial}{\partial y} f(x, y) = \frac{\partial^2}{\partial y^2} f(x, y)$	$f_{yy}(x, y)$
$\frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y) = \frac{\partial^2}{\partial x \partial y} f(x, y)$	$f_{xy}(x, y)$
$\frac{\partial}{\partial y} \frac{\partial}{\partial x} f(x, y) = \frac{\partial^2}{\partial y \partial x} f(x, y)$	$f_{yx}(x, y)$

Example 25.3. Find the four second-order partial derivatives of $f(x, y) = 3x^2y^3 + 4xy - 3x^2$

Note that the order in which mixed partial derivatives are taken can sometimes change the result. However, it is often the case that the order of the partial derivatives can be switched.

Theorem 25.1. *Let the function $f(x, y)$ be defined on an open domain \mathcal{D} of \mathbb{R}^2 and assume that f_{xy} and f_{yx} are continuous over the domain \mathcal{D} . Then, $f_{xy} = f_{yx}$ for all points in the domain \mathcal{D} .*

Many of the commonly used functional forms in data science meet the criteria above. Thus, for many of the commonly used functions in data science, the order of evaluation of partial derivatives often does not matter. In practice, it is always good practice to verify this though.

Chapter 26

The chain rule

```
library(tidyverse)
library(plotly)
library(dasc2594)
set.seed(2021)
```

Recall the univariate chain rule: If $y = f(x)$ is a function of x and $z = g(y)$ is a function of y , a question of interest is "What is the change in z relative to change in x ?

We can write $z = g(y) = g(f(x))$ and using this notation, the change in z with respect to the variable x is $\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = \frac{df(y)}{dy} \frac{dg(x)}{dx}$. Written in functional form

$$(g(f(x)))' = (g \circ f)'(x) = g'(f(x))f'(x)$$

Example 26.1. Let $z = f(y) = y^3$ and let $y = g(x) = e^x$ what is $\frac{dz}{dx}$?

26.1 The chain rule with one independent variable

Drawing in class

Definition 26.1 (Chain rule for one independent variable). Let z be a differentiable function of two variables x and y so that $z = f(x, y)$ and let $x = g(t)$ be a function of t and $y = h(t)$ a function of t . Written in functional form, z can be written as $z = f(x, y) = f(g(t), h(t))$, with $x = g(t)$ and $y = h(t)$. Then we can define the derivative of z with respect to t as

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt}$$

- For the definition above, we have the dependent variable z and we have **intermediate variables** x and y .
- Notice in the definition above that there is a mix of partial derivatives (∂) and ordinary derivatives (d).

Example 26.2. Let $z = x^2 + e^y$ and let $x = \cos(t)$ and $y = \sin(t)$

The results from above can also be extended to have more than two intermediate variables.

Draw picture

26.2 The chain rule with several independent variables

Often, functions will have more than one independent variables.

Definition 26.2 (Chain rule for two independent variables). Let z be a differentiable function of two variables x and y so that $z = f(x, y)$ and let $x = g(t, s)$ be a function of s and t and $y = h(s, t)$ a function of s and t . Written in functional form, z can be written as $z = f(x, y) = f(g(s, t), h(s, t))$, with $x = g(s, t)$ and $y = h(s, t)$. Then we can define the partial derivative of z with respect to s as

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s}$$

the partial derivative of z with respect to t as

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t}$$

Example 26.3. Let $z = f(x, y) = x^2 e^y$ and let $x = 2s - t$ and $y = 4s^3 - 3t^2$

26.3 The chain rule in matrix notation

To get a better understanding of the chain rule, it helps to show the chain rule using matrix notation. Using the matrix notation will enable you to apply the chain rule to any number of intermediate variables. For example, consider the extension of the definition for the chain rule of a function with one independent variable.

Definition 26.3 (Matrix chain rule for one independent variable). Let z be a differentiable function of two variables x and y so that $z = f(x, y)$ and let $x = g(t)$ be a function of t and $y = h(t)$ a function of t . Written in functional

form, z can be written as $z = f(x, y) = f(g(t), h(t))$, with $x = g(t)$ and $y = h(t)$. Then we can define the derivative of z with respect to t as

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt}$$

Written in matrix notation, this is

$$\frac{dz}{dt} = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt}$$

The definition above for the chain rule with two variables is given by

Definition 26.4 (Chain rule for two independent variables). Let z be a differentiable function of two variables x and y so that $z = f(x, y)$ and let $x = g(s, t)$ be a function of s and t and $y = h(s, t)$ a function of s and t . Written in functional form, z can be written as $z = f(x, y) = f(g(s, t), h(s, t))$, with $x = g(s, t)$ and $y = h(s, t)$. Then we can define the partial derivative of z with respect to s as

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s}$$

which, in matrix notation is

$$\frac{dz}{dt} = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial s} \end{pmatrix} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s}$$

The partial derivative of z with respect to t as

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t}$$

which, in matrix notation is

$$\frac{dz}{dt} = \begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial t} \end{pmatrix} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t}$$

This use of matrix notation for derivatives will be useful in understanding the **gradient**.

Chapter 27

The gradient and directional derivatives

```
library(tidyverse)
library(plotly)
library(dasc2594)
set.seed(2021)
```

Partial derivatives tell about how a rate of function changes in a particular direction (in the direction of a coordinate).

Think about trying to find the maximum of a real-valued function (finding the minimum is equivalent to finding the maximum of the negative value of the function). Finding the maximum of a function is analogous to hiking up a mountain and trying to find the highest peak.

Suppose you are standing on a mountain surface at the point (x, y, z) in 3-dimensions where $z = f(x, y)$ is the function that gives the height of the mountain at location (x, y) . If you are standing at the point (a, b) in the (x, y) coordinate system, you might want to get to the top of the mountain as quickly as possible. The direction that is the steepest uphill direction can be calculated using the concepts of the **directional derivative** and the **gradient**.

Definition 27.1 (Directional Derivative). Given a function $f(x, y)$ that is differentiable at (a, b) and a unit vector $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ in the $x - y$ plane, the **directional derivative** of f at (a, b) in the direction of \mathbf{v} is

$$D_{\mathbf{v}}f(a, b) = \lim_{h \rightarrow 0} \frac{f(a + hv_1, b + hv_2) - f(a, b)}{h}$$

assuming the limit exists.

To understand how the directional derivative relates to partial derivatives, in the definition above, let $v_2 = 0$ and to make \mathbf{v} a unit vector, set $v_1 = 1$ (\mathbf{v} is the standard basis vector \mathbf{e}_1). Then, the limit in the directional derivative definition above becomes

$$\lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h}$$

which is the definition for the partial derivative of f in the x direction $f_x = \frac{\partial f}{\partial x}$ (Definition 25.1). Likewise, letting $\mathbf{v} = \mathbf{e}_2$, the standard basis vector in the y direction, gives the partial derivative in the y direction $f_y = \frac{\partial f}{\partial y}$. Thus, one could pick any direction vector \mathbf{v} and then calculate the partial derivative in that direction.

Now observe that any line that goes through the point (a, b) in the direction of the unit vector $\mathbf{v} \in \mathcal{R}^2$ can be written as the set of points $\{(x = a + sv_1, y = b + sv_2) | s \in \mathcal{R}\}$ which forms a line through the point (a, b) in the direction of \mathbf{v} . In this definition, the value s determines the length of the vector because \mathbf{v} is a unit vector. At $s = 0$, this definition corresponds to the point (a, b) and as s increases, the points (x, y) are the set of points along the line that are distance $|s|$ away from (a, b) . Notice that this set defines a function $g(s) = f(a + sv_1, b + sv_2) = f(x, y)$ which is a single variable function of the two inputs x and y of $f(x, y)$. Given this definition, the directional derivative of $f(x, y)$ in the direction of \mathbf{v} at the point (a, b) is now given by

$$\begin{aligned} D_{\mathbf{v}} f(a, b) &= \frac{d}{ds} g(s)|_{s=0} \\ &= \frac{\partial f}{\partial x} \frac{dx}{ds} + \frac{\partial f}{\partial y} \frac{dy}{ds}|_{s=0} \\ &= f_x(a, b)v_1 + f_y(a, b)v_2 \\ &= (f_x(a, b) \quad f_y(a, b)) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \end{aligned}$$

which is the dot product of the vectors $\begin{pmatrix} f_x(a, b) \\ f_y(a, b) \end{pmatrix}$ and \mathbf{v} .

Notice that the vector \mathbf{v} is a unit vector and therefore the directional derivative is a weighted sum of the partial derivatives in the x and y directions weighted by the vector \mathbf{v} (weighted sums are sums where the coefficients sum to 1—in this case the sum is in the “distance” metric). As a consequence, we can find the directional derivative in any direction by changing the vector \mathbf{v} .

Definition 27.2. Let $f(x, y)$ be a differentiable function at (a, b) and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ a unit vector in the xy plane. Then, the directional derivative of f at (a, b) in the direction of \mathbf{v} is

$$D_{\mathbf{v}} f(a, b) = \left(\frac{\partial f(x, y)}{\partial x} \Big|_{(x, y)=(a, b)} \quad \frac{\partial f(x, y)}{\partial y} \Big|_{(x, y)=(a, b)} \right) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Example 27.1. Compute the directional derivative of $f(x, y) = 3x^2 + y^2$ in the direction of $\mathbf{u} = \left(\frac{1}{\sqrt{3}}, \frac{\sqrt{2}}{\sqrt{3}}\right)$ and $\mathbf{v} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ at the point $(2, 1)$.

- calculate the directional derivatives
- graph the directional derivatives using contour plots and segments

```
# the function
target_fun <- function(x, y) {
  3 * x^2 + y^2
}

# the gradient function
gradient_fun <- function(x, y) {
  c(6*x, 2 * y)
}

# define the unit vectors u and v
u <- c(1 / sqrt(3), sqrt(2/3))
v <- c(1 / sqrt(2), -1 / sqrt(2))

# create a set of gridpoints for plotting the function
N <- 50
x <- seq(-10, 10, length.out = N)
y <- seq(-10, 10, length.out = N)

dat <- expand_grid(x = x, y = y) %>%
  mutate(z = target_fun(x, y))

# define the point (a, b)
a <- 2
b <- 1

# directional derivative of f in the direction of u
Du_ab <- sum(gradient_fun(a, b) * u)
Du_ab

## [1] 8.561196

Dv_ab <- sum(gradient_fun(a, b) * v)
Dv_ab

## [1] 7.071068
```

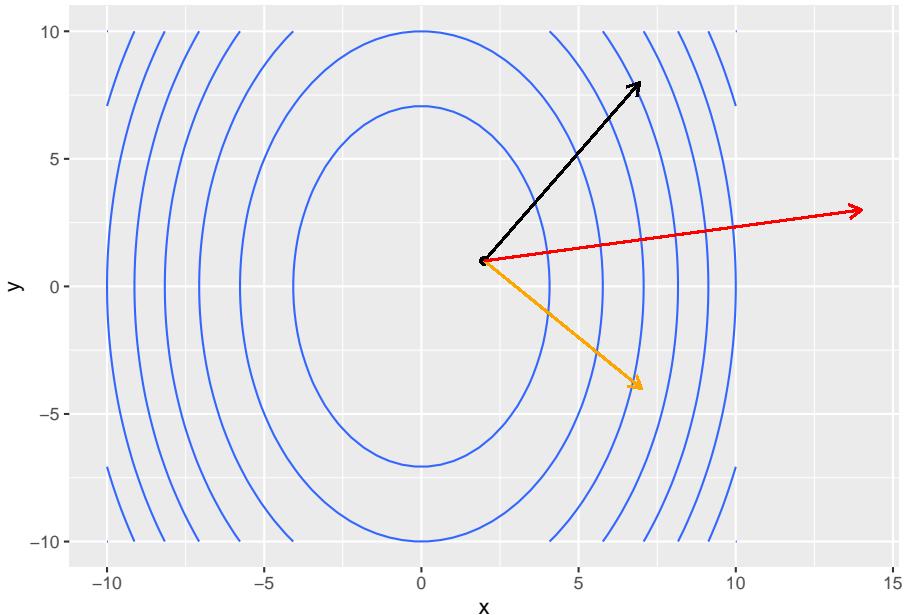
```

q <- c(1, 0)
Dq_ab <- sum(gradient_fun(a, b) * q)
Dq_ab

## [1] 12

# generate the plot
ggplot(dat, aes(x = x, y = y, z = z)) +
  geom_contour() +
  geom_point(aes(x = a, y = b)) +
  geom_segment(aes(x = a, y = b, xend = a + u[1] * Du_ab, yend = b + u[2] * Du_ab,
                  arrow = arrow(length = unit(0.1, "in")))) +
  geom_segment(aes(x = a, y = b, xend = a + v[1] * Dv_ab, yend = b + v[2] * Dv_ab,
                  arrow = arrow(length = unit(0.1, "in"))), color = "orange") +
  geom_segment(aes(x = a, y = b, xend = a + gradient_fun(a, b)[1], yend = b + gradient_fun(a, b)[2],
                  arrow = arrow(length = unit(0.1, "in"))), color = "red")

```



27.1 The Gradient

The directional derivative is a dot product of the partial derivatives and a unit vector. The gradient is similar, but rather than return a single value (a number), the gradient returns a vector at a point (a, b) .

Definition 27.3 (The Gradient). Let $f(x, y)$ be a differentiable function at (a, b) . Then, the gradient of f at (a, b) is

$$\begin{aligned}\nabla f(a, b) &= \left(\frac{\partial f(x, y)}{\partial x} \Big|_{(x, y)=(a, b)}, \frac{\partial f(x, y)}{\partial y} \Big|_{(x, y)=(a, b)} \right) \\ &= \frac{\partial f(x, y)}{\partial x} \Big|_{(x, y)=(a, b)} \mathbf{e}_1 + \frac{\partial f(x, y)}{\partial y} \Big|_{(x, y)=(a, b)} \mathbf{e}_2,\end{aligned}$$

where $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are the standard basis vectors in \mathbb{R}^2 .

Notice that the directional derivative at the point (a, b) can be calculated using the gradient where

$$\begin{aligned}D_{\mathbf{v}} f(a, b) &= \nabla f(a, b) \cdot \mathbf{v} \\ &= \left(\frac{\partial f(x, y)}{\partial x} \Big|_{(x, y)=(a, b)}, \frac{\partial f(x, y)}{\partial y} \Big|_{(x, y)=(a, b)} \right) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\end{aligned}$$

the directional derivative is the dot product of the gradient $\nabla f(a, b)$ at the point (a, b) with the unit vector \mathbf{v} .

Example 27.2. Compute the gradient of $f(x, y) = 3x^2 + y^2$ at the point $(3, 1)$.

- calculate the gradient
- graph the gradient using contour plots and segments

The gradient is critical in data science because is the tool that allows for finding the set of parameters for a given model that are “most likely” given the data. The gradient has the property in that at each point (a, b) where $f(x, y)$ is differentiable, the gradient points in the direction of the maximum rate of change.

Theorem 27.1 (The gradient and rates of change). *Let $f(x, y)$ be a differentiable function at (a, b) with $\nabla f(a, b) \neq 0$. Then,*

- 1) *f has its maximum rate of increase at the point (a, b) in the direction of the gradient $\nabla f(a, b)$. Because the gradient is a weighted sum of the partial derivatives and the unit vector in the direction of the maximum change, the magnitude of the rate of change is $\|\nabla f(a, b)\|$ which is the length of the gradient vector.*
- 2) *f has its maximum rate of decrease at the point (a, b) in the direction of the gradient $-\nabla f(a, b)$. The rate of change in the direction of maximum rate of decrease is $-\|\nabla f(a, b)\|$.*
- 3) *The directional derivative is 0 in any direction orthogonal to $\nabla f(a, b)$.*

Example 27.3. Consider the function $f(x, y) = 3x^2 - 2xy + y^2$. At the point $(3, 1)$, what is the direction of steepest descent? Steepest ascent?

- graph the function as contours and plot the gradient as a segment

27.1.1 The gradient and the tangent line

Theorem 27.1 states that the directional derivative at the point (a, b) is 0 in any direction that is orthogonal to $\nabla f(a, b)$. Because the directional derivative is the rate of change of the function in the direction of \mathbf{v} , the directional derivative being 0 means that the function $f(x, y)$ is not changing in the direction of the vector \mathbf{v} . Therefore, we know that the vector \mathbf{v} and the vector $\nabla f(x, y)|_{(a,b)}$ are orthogonal because the definition of the directional derivative in definition 27.2 states that

$$D_{\mathbf{v}}f(a, b) = \nabla f(a, b) \cdot \mathbf{v} = (\nabla f(a, b))' \mathbf{v} = 0$$

which is only true if the gradient $\nabla f(a, b)$ is orthogonal to \mathbf{v} . Because the vector \mathbf{v} points in the direction of 0 change in $f(x, y)$, the vector \mathbf{v} is a tangent line to the level curve. **See drawing**

Using this, one can calculate the tangent to the level curve at the point (a, b) as the dot-product equation

$$(f_x(x, y) \quad f_y(x, y)) \begin{pmatrix} x - a \\ y - b \end{pmatrix} = f_x(x - a) + f_y(y - b) = 0.$$

Example 27.4. For the function $f(x, y) = 3x^2 - 2xy + y^3$, find a vector orthogonal to the gradient $\nabla f(a, b)$ at the point $(a, b) = (2, 1)$.

27.1.2 The gradient in higher dimensions

We can extend the gradient to higher dimensional functions. Let $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_n)'$ be a vector in \mathcal{R}^n . Then the gradient of f at a point $\mathbf{a} = (a_1, a_2, \dots, a_n)$ is

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1}|_{\mathbf{x}=\mathbf{a}} \\ \frac{\partial f(\mathbf{x})}{\partial x_2}|_{\mathbf{x}=\mathbf{a}} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n}|_{\mathbf{x}=\mathbf{a}} \end{pmatrix} = \begin{pmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{pmatrix}$$

Chapter 28

Tangent planes and linear approximations

```
library(tidyverse)
library(plotly)
library(dasc2594)
set.seed(2021)
```

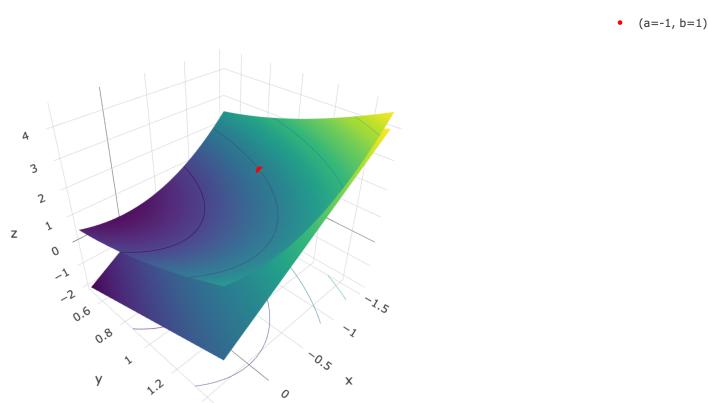
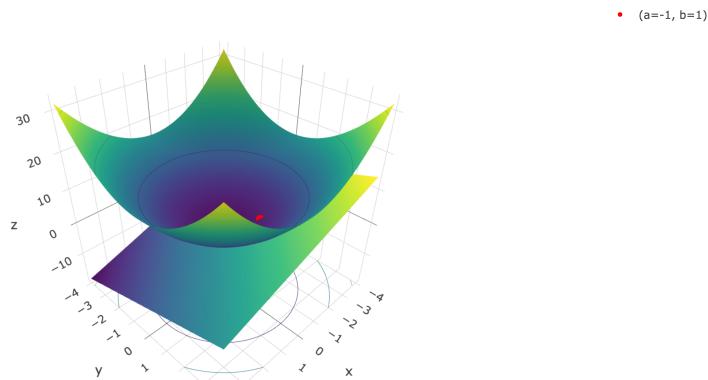
Let $f(\mathbf{x})$ be a differentiable function at a point \mathbf{a} . Because the function is differentiable at \mathbf{a} , this means that all paths $P_{\mathbf{a}}$ that approach the point \mathbf{a} from all directions all take on values $f(P_{\mathbf{a}})$ that are “close” to $f(\mathbf{a})$. Mathematically, we describe this as **smoothness**. A more explicit description says that as the paths $P_{\mathbf{a}}$ get very close to \mathbf{a} , the space over which these paths are defined starts to look more and more like a flat surface—the **tangent plane**.

Example 28.1. For this example, we plot the function $f(x, y) = x^2 + y^2$ which has gradient $\nabla f(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$

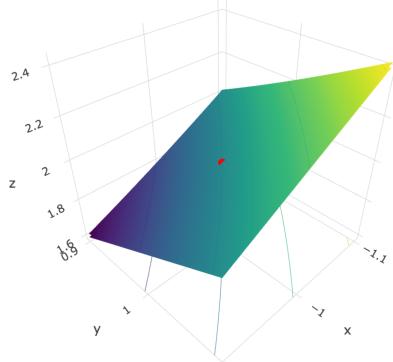
```
# f(x, y)
target_fun <- function(x, y) {
  return(x^2 + y^2)
}
# gradient f(x, y)
grad_fun <- function(x, y) {
  c(2 * x, 2 * y) # notice that the return value is a vector
}
# plot
plot_tangent_plane(target_fun = target_fun, grad_fun = grad_fun, a=-1, b = 1)
```

230 CHAPTER 28. TANGENT PLANES AND LINEAR APPROXIMATIONS

```
# zoomed in plot
plot_tangent_plane(target_fun = target_fun, grad_fun = grad_fun, a=-1, b = 1, xlim = c
# super zoomed in plot
plot_tangent_plane(target_fun = target_fun, grad_fun = grad_fun, a=-1, b = 1, xlim = c
```



• (a=-1, b=1)



A consequence of this result that when you zoom in on a differentiable function the function looks like a flat plane is that the function $f(x, y)$ can be approximated locally as a linear function (local approximation just means that if you are really “close” to the point (a, b) that the function will behave like a tangent plane if the function is differentiable). Intuitively, this makes sense as if the derivative exists, the directional derivatives are just vectors and a linear combination of vectors (in \mathbb{R}^2) produces a tangent plane (in higher dimensions, this is called a **hyperplane**). This means that if the function $f(x, y)$ is differentiable at the point (a, b) , then $f(x, y)$ for points (x, y) close to (a, b) is approximated by the linear tangent plane.

Notice in the code above that there are two functions needed to calculate the tangent plane: the function $f(x, y)$ and the gradient $\nabla f(x, y)$. This can be seen in the definition of the tangent plane.

Definition 28.1 (Tangent plane). Let $f(x, y)$ be a differentiable function at the point (a, b) . Then the tangent plane to the surface defined by the function $f(x, y)$ at the point (a, b) is given by the equation

$$\begin{aligned} z &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &= f(a, b) + \nabla f(x, y)|_{(a,b)} \cdot \begin{pmatrix} x - a \\ y - b \end{pmatrix} \\ &= f(a, b) + (\nabla f(x, y)|_{(a,b)})' \begin{pmatrix} x - a \\ y - b \end{pmatrix} \end{aligned}$$

,

where the tangent plane is defined as the dot product of the gradient vector $\nabla f(x, y)|_{(a,b)}$ evaluated at the point (a, b) and the vector $\begin{pmatrix} x - a \\ y - b \end{pmatrix}$ that is the coordinate-wise distance of the point (x, y) from the point (a, b) .

Example 28.2. Find the equation for the tangent plane for the function $f(x, y) = x^2 \cos(y) - y^2 \cos(x)$ at the point $(\frac{\pi}{2}, \frac{pi}{4})$

- calculate by hand
- plot using `plot_tangent_plane()`

This leads to the linearization equation for functions of n variables.

Definition 28.2 (Linearization of a function). Let $f(\mathbf{x})$ be a differentiable function at the point $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ for a function of inputs $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in \mathcal{R}^n$. Then the linearization of the function $f(\mathbf{x})$ at the point \mathbf{a} is given by the equation

$$\begin{aligned} L(\mathbf{x}) &= f(\mathbf{a}) + \nabla f(\mathbf{x})|_{\mathbf{a}} \cdot (\mathbf{x} - \mathbf{a}) \\ &= f(\mathbf{a}) + (\nabla f(\mathbf{x})|_{\mathbf{a}})' (\mathbf{x} - \mathbf{a}) \end{aligned}$$

The quality of the linearization is high for points “close” to \mathbf{a} and has higher error (defined as $\|L(\mathbf{x}) - f(\mathbf{x})\|$) as \mathbf{x} gets further from \mathbf{a} .

For points \mathbf{x} “close” to the point \mathbf{a} , the exact difference in the function $z = f(\mathbf{x})$ is given by $\Delta z = f(\mathbf{x}) - f(\mathbf{a})$. Plugging in the linear approximation, the differential $dz = L(\mathbf{x}) - f(\mathbf{a})$ is the linear approximation to the exact different

Δz . Define $d\mathbf{x} = \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix} = \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{pmatrix}$ as the set of changes in each of the n

coordinates with respect to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ so that the linear change dz of $f(\mathbf{x})$ at \mathbf{a} is given by

$$\begin{aligned} dz &= \nabla f(\mathbf{x})|_{\mathbf{a}} \cdot (\mathbf{x} - \mathbf{a}) \\ &= \nabla f(\mathbf{x})|_{\mathbf{a}} \cdot d\mathbf{x} \\ &= \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} dx_i, \end{aligned}$$

where the last term is a sum of the linear approximation in each of the $i = 1, \dots, n$ coordinate directions.

Example 28.3. Approximate the linear change of the function $f(x, y, z) = x^2 - 3xy^2z^2 - 4z^2$ at the point $(a, b, c) = (1, -2, -1)$ evaluated at the point $(x, y, z) = (0.95, -2.05, -1.05)$. Compare this to the exact value of the function $f(x, y, z)$

First, we find the gradient

$$\nabla f(x, y, z) = \begin{pmatrix} 2x - 3y^2z^2 \\ -6xyz^2 \\ -6xy^2z - 8z \end{pmatrix}$$

and evaluate the gradient at $(a, b, c) = (1, -2, -1)$ to get

$$\nabla f(1, -2, -1) = \begin{pmatrix} 2(1) - 3(-2)^2(-1)^2 \\ -6(1)(-2)(-1)^2 \\ -6(1)(-2)^2(-1) - 8(-1) \end{pmatrix} = \begin{pmatrix} -10 \\ 12 \\ 32 \end{pmatrix}$$

The function evaluated at the point (a, b, c) is $f(1, -2, -1) = (1)^2 - 3(1)(-2)^2(-1)^2 - 4(-1)^2 = -15$.

Therefore, the linear approximation $L(x, y, z)$ of $f(x, y, z)$ at the point (a, b, c) is

$$\begin{aligned} L(x, y, z) &= f(a, b, c) + \nabla f(a, b, c) \cdot \begin{pmatrix} x - 1 \\ y - (-2) \\ z - (-1) \end{pmatrix} \\ &= -15 + 10(x - 1) - 12(y + 2) - 32(z + 1) \end{aligned}$$

The linear approximation evaluated at the point $(0.95, -2.05, -1.05)$ is

$$L(0.95, -2.05, -1.05) = -15 - 10(0.95 - 1) + 12(-2.05 + 2) + 32(-1.05 + 1) = -16.7.$$

Compared to the true value of the function $f(x, y, z)$ is

$$f(0.95, -2.05, -1.05) = (1.05)^2 - 3(1.05)(-2.05)^2(-1.05)^2 - 4(-1.05)^2 = -16.71228,$$

which gives an approximation error of $f(x, y, z) - L(x, y, z) = -16.7122803 - -16.7 = -0.0122803$

```
target_fun <- function(x, y, z) {
  x^2 - 3 * x * y^2 * z^2 - 4 * z^2
}

grad_fun <- function(x, y, z) {
  c(2*x - 3 * y^2 * z^2,
    -6 * x * y * z^2,
    -6 * x * y^2 * z - 8 * z)
}

grad_fun(1, -2, -1)
```

```
## [1] -10 12 32
```

```
target_fun(0.95, -2.05, -1.05)

## [1] -16.71228

linearization <- function(target_fun, grad_fun, x, y, z, a, b, c) {
  target_fun(a, b, c) + sum(grad_fun(a, b, c) * c(x - a, y - b, z - c))
}

linearization(target_fun, grad_fun, 0.95, -2.05, -1.05, 1, -2, -1)

## [1] -16.7

target_fun(0.95, -2.05, -1.05)

## [1] -16.71228

# approximation error
target_fun(0.95, -2.05, -1.05) - linearization(target_fun, grad_fun, 0.95, -2.05, -1.05)
```

```
## [1] -0.01228031
```

Chapter 29

Minimums and Maximums

```
library(tidyverse)
library(plotly)
library(dasc2594)
set.seed(2021)
```

In general, we talk about finding either minimums or maximum values of a function. For simplicity, we focus here on finding the minimum value of a function $f(\cdot)$ because finding the maximum value of $f(\cdot)$ is equivalent to finding the minimum value of $-f(\cdot)$.

To characterize these minimum values, we consider two different types of minimums: local minimums and global minimums. For now, we focus on functions of two variables but these ideas are similar for functions of many variables.

29.1 Local minimums

A local minimum of the function $f(x, y)$ is a point (a, b) where the values of the function $f(a, b)$ evaluated at (a, b) is less than or equal to the function $f(a + \Delta x, b + \Delta y)$ at any nearby point $(a + \Delta x, b + \Delta y)$, where Δx and Δy are very small values.

Definition 29.1. Let (a, b) be a point in the domain \mathcal{D} of the function $f(x, y)$. If there exists an $\epsilon > 0$ such that $\|(x, y) - (a, b)\| < \epsilon$ (the point (x, y) is close to the point (a, b) for a given disk of radius ϵ at the point (a, b)), then if $f(a, b) \leq f(x, y)$, the point (a, b) is called a **local minimum** value of the function $f(x, y)$.

In terms of finding a minimum of a surface, the local minimum is any point on the surface from which one cannot walk downhill if one can only take small steps. Like in univariate functions, local minimums also have a relationship to the partial derivatives.

Theorem 29.1. *If $f(x, y)$ has a local minimum at the point (a, b) and $f(x, y)$ is a differentiable function at (a, b) , then the partial derivatives $\frac{\partial f(x, y)}{\partial x}|_{(a, b)} = f_x(a, b) = 0$ and $\frac{\partial f(x, y)}{\partial y}|_{(a, b)} = f_y(a, b) = 0$*

Note: the converse is not necessarily true: just because $f_x(a, b) = 0$ and $f_y(a, b) = 0$, this doesn't mean that we have a minimum point at (a, b) .

As stated above, the fact that the partial derivatives $f_x(a, b) = 0$ and $f_y(a, b) = 0$ does not imply that the point (a, b) is a local minimum. Instead, the partial derivatives being 0 only implies that (a, b) is a critical point of the function $f(x, y)$.

Definition 29.2. The point (a, b) is a critical point of the function $f(x, y)$ if either

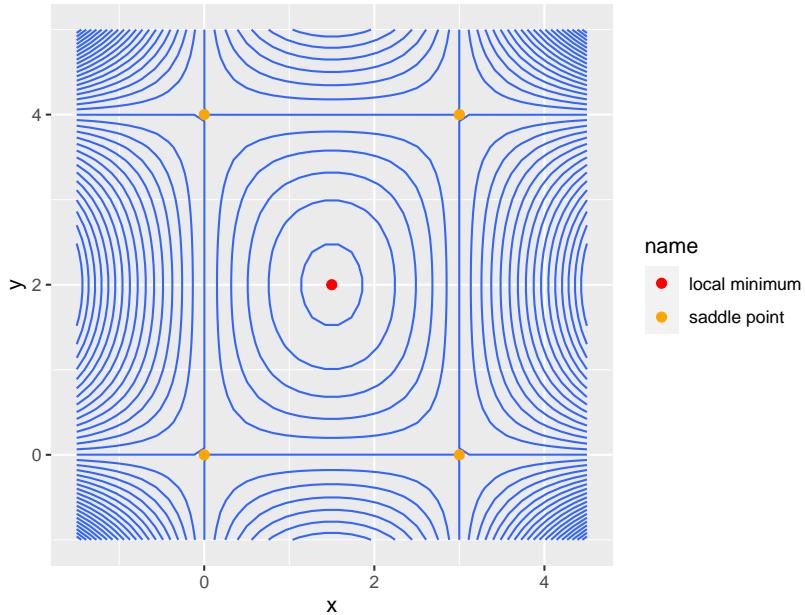
- a) $f_x(a, b) = 0$ and $f_y(a, b) = 0$ or
- b) at least one partial derivative does not exist at (a, b) .

Example 29.1. find the critical points of $f(x, y) = xy(x - 3)(y - 4)$. * plot the critical points using `plotly()`

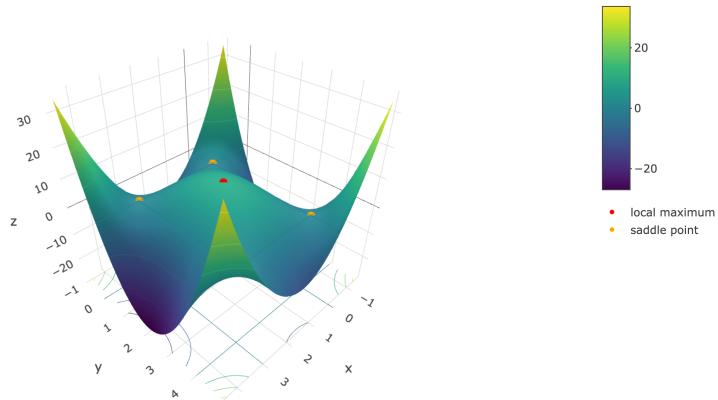
```
f <- function(x, y) {
  x * y * (x - 3) * (y - 4)
}
x <- seq(3/2 + -3, 3/2 + 3, length.out = 40)
y <- seq(2 + -3, 2 + 3, length.out = 40)

critical_points <- data.frame(x = c(3/2, 0, 3, 0, 3), y = c(2, 0, 0, 4, 4)) %>%
  mutate(z = f(x, y), color = c("red", rep("orange", 4)), name = c("local minimum", 
  dat <- expand_grid(x, y) %>%
  mutate(z = f(x, y))

dat %>%
  ggplot(aes(x = x, y = y, z = z)) +
  geom_contour(bins = 40) +
  coord_fixed(ratio = 1) +
  geom_point(data = critical_points, aes(color = name), size = 2) +
  scale_color_manual(values = critical_points$color)
```



```
plot_ly(x = x, y = y, z = matrix(dat$z, 40, 40)) %>%
  add_surface(
    contours = list(
      z = list(
        show=TRUE,
        usecolormap=TRUE,
        highlightcolor="#ff0000",
        project=list(z=TRUE)))) %>%
  add_trace(x = critical_points$x, y = critical_points$y,
            z = critical_points$z,
            mode = "markers", type = "scatter3d",
            marker = list(size = 5,
                          color = c("red", rep("orange", 4)),
                          symbol = 104),
            name = c("local maximum", rep("saddle point", 4)))
```

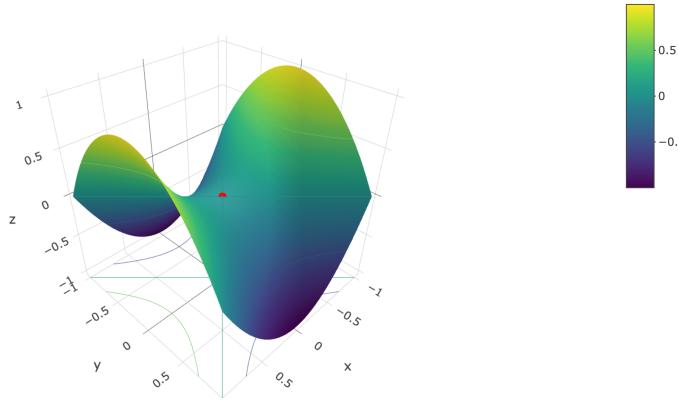


If a critical point isn't a local minimum/maximum, what else could this point be? One possible shape is what is called a saddle point. A saddle point is one in which the function $f(x, y)$ is increasing in one direction (say, along the x-axis) and decreasing along another direction (say, the y-axis). The surface is often said to resemble the shape of a saddle (or a Pringles chip). An example saddle point is shown in the plot below.

```
f <- function(x, y) {
  x^2 - y^2
}
x <- seq(-1, 1, length.out = 40)
y <- seq(-1, 1, length.out = 40)
critical_points <- data.frame(x = 0, y = 0) %>%
  mutate(z = f(x, y))
dat <- expand_grid(x, y) %>%
  mutate(z = f(x, y))

plot_ly(x = x, y = y, z = matrix(dat$z, 40, 40)) %>%
  add_surface(
    contours = list(
      z = list(
        show=TRUE,
        usecolormap=TRUE,
        highlightcolor="#ff0000",
        project=list(z=TRUE)))) %>%
  add_trace(x = critical_points$x, y = critical_points$y,
            z = critical_points$z,
            mode = "markers", type = "scatter3d",
            marker = list(size = 5,
                          color = "red",
                          opacity = 0.8),
            name = "Critical Points")
```

```
symbol = 104),
name = "saddle point")
```



Definition 29.3 (Saddle point). A saddle point of a differentiable function $f(x, y)$ is a critical point (a, b) of $f(x, y)$ where for all $\epsilon > 0$ there exists points (x, y) within the disk of radius ϵ of the point (a, b) (i.e., $\|(x, y) - (a, b)\| < \epsilon$) where $f(x, y) > f(a, b)$ and where $f(x, y) < f(a, b)$. In other words, there are points nearby the critical point that have both higher and lower values than the value $f(a, b)$.

Finding critical points is a first step in finding minimums, however, we need a method to determine whether a critical point is a local minimum/maximum or a saddle point. To determine this, the second derivative test is useful.

Theorem 29.2 (Second Derivative Test). Let $f_x(a, b) = f_y(a, b) = 0$ so that the point (a, b) is a critical point. Also let the $f(x, y)$ be second-order differentiable (the second partial derivatives exists) throughout the open disk $\|(x, y) - (a, b)\| < \epsilon$ for some $\epsilon > 0$. Then

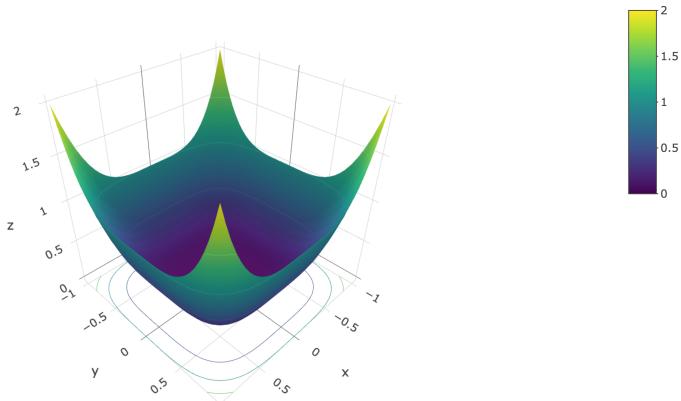
- 1) If $f_{xx}(a, b)f_{yy}(a, b) - (f_{xy}(a, b))^2 > 0$ and $f_{xx}(a, b) < 0$, then $f(x, y)$ has a local maximum at (a, b) .
- 2) If $f_{xx}(a, b)f_{yy}(a, b) - (f_{xy}(a, b))^2 > 0$ and $f_{xx}(a, b) > 0$, then $f(x, y)$ has a local minimum at (a, b) .
- 3) If $f_{xx}(a, b)f_{yy}(a, b) - (f_{xy}(a, b))^2 < 0$, then $f(x, y)$ has a saddle point at (a, b) .
- 4) If $f_{xx}(a, b)f_{yy}(a, b) - (f_{xy}(a, b))^2 = 0$, the test is inconclusive.

Example 29.2. find the critical points of $f(x, y) = xy(x - 3)(y - 4)$ and determine if they are local maxima, minima, saddle points, or undetermined.

Example 29.3. find the critical points of $f(x, y) = x^4 + y^4$ and determine if they are local maxima, minima, saddle points, or undetermined.

```
f <- function(x, y) {
  x^4 + y^4
}
x <- seq(-1, 1, length.out = 40)
y <- seq(-1, 1, length.out = 40)
critical_points <- data.frame(x = 0, y = 0) %>%
  mutate(z = f(x, y))
dat <- expand_grid(x, y) %>%
  mutate(z = f(x, y))

plot_ly(x = x, y = y, z = matrix(dat$z, 40, 40)) %>%
  add_surface(
    contours = list(
      z = list(
        show=TRUE,
        usecolormap=TRUE,
        highlightcolor="#ff0000",
        project=list(z=TRUE)))) %>%
  add_trace(x = critical_points$x, y = critical_points$y,
            z = critical_points$z,
            mode = "markers", type = "scatter3d",
            marker = list(size = 5,
                          color = "red",
                          symbol = 104),
            name = "saddle point")
```



29.2 Global maximums and minimums

We have shown how you can find critical points (a, b) by finding points that have 0 partial derivatives in each direction ($f_x(a, b) = f_y(a, b) = 0$) and we have shown how to characterize these critical points using the second derivative test. In data science, we are often interested in finding the “best” model, not just the “best local” model. So a question arises is whether the local minima/maxima that we find are the global minima/maxima. To determine this, we have to define the global minima/maxima.

Definition 29.4. If $f(x, y)$ is a function defined over a domain (or a subset of the domain) $\mathcal{D} \in \mathbb{R}^2$ that contains the point (a, b) , we say

- (a, b) is a **global maximum** if $f(a, b) \geq f(x, y)$ for all $(x, y) \in \mathcal{D}$ and
- (a, b) is a **global minimum** if $f(a, b) \leq f(x, y)$ for all $(x, y) \in \mathcal{D}$.

Thus, once you have found all the critical points, you also need to check all the boundary points for the domain (or subdomain) \mathcal{D} as the global minimum/maximum might lie on the boundary.

In general finding the global maximum/minimum points can be done by applying the following procedure:

Let $f(x, y)$ be a function on a closed (the set contains its boundary points) and bounded (the set is not infinite) set. Then, the global minima/maxima can be found by

- 1) Determining the values $f(x, y)$ at all the critical points.
- 2) Determining the values of $f(x, y)$ at all the boundary points of \mathcal{D} .
- 3) Finding the minimum/maximum of the set of function values from 1) and 2).

Example 29.4. Let $f(x, y) = x^2 - 4xy + y^2$ over the region $[-2, 2] \times [-2, 2]$

In data science problems, it is often difficult to analyze the functions analytically (using the tools of calculus directly). Instead, one can use numeric techniques to find minima and maxima. Two techniques that we will discuss are grid search and gradient descent.

29.3 Grid search optimization

If the number of input variables in the function is small, one can find global minima and maxima using grid search. To find a global minima/maxima in \mathbb{R}^2 , one creates a grid over the domain \mathcal{D} and evaluates the function $f(x, y)$ at these grid points and finds the numerical solution.

Example 29.5. Let $f(x, y) = x^2 - 2xy + y^2$ over the region $[-2, 2] \times [-2, 2]$

```
target_fun <- function(x, y) {
  x^2 - 4 * x * y + y^2
}
# N is the grid size
N <- 1000
x <- seq(-2, 2, length.out = N)
y <- seq(-2, 2, length.out = N)
dat <- expand_grid(x = x, y = y) %>%
  mutate(z = target_fun(x, y))
# find the minimum
dat %>%
  filter(z == min(z))

## # A tibble: 2 x 3
##       x     y     z
##   <dbl> <dbl> <dbl>
## 1    -2    -2    -8
## 2     2     2    -8

# find the maximum
dat %>%
  filter(z == max(z))

## # A tibble: 2 x 3
##       x     y     z
##   <dbl> <dbl> <dbl>
## 1    -2     2    24
## 2     2    -2    24
```

29.4 Gradient Descent

Another method to find minima/maxima is gradient descent.