A short introduction to university algebra

22 lectures for online study during Covid-19

Graham Ellis

PUBLISHED BY *Shoe String Press*

HTTP://HAMILTON.NUIGALWAY.IE

First release, October-December 2020

Contents

III Eigenvalues

Arithmetic

1. Basic arithmetic

What might we say about the following four equations?

$$5+4=9 \tag{1.1}$$

$$10+11=9 \tag{1.2}$$

$$298+71=9 \tag{1.3}$$

$$9=9 \tag{1.4}$$

We might be inclined to say the first is correct, the second and third are incorrect, and the last is obvious. Certainly, if Jessie has a basket of 5 apples, and she adds 4 apples to it, then she'll have a basket of 9 apples. If she has a basket of 10 apples, and adds 11 more to it, then she'll not have a basket of 9 apples. On the other hand, if Mary starts work at 10 o'clock, and she works for 11 hours, then she'll finish work at 9 o'clock. If Joseph is sailing on a course of 298°, and he decides to add 71° to his course, then he'll be on a course of 9°. So statements (1.2) and (1.3) can be true, and it's a bit rash to declare them outright incorrect..

Maybe it is safer to say that equation (1.1) is always true, equations (1.2) and (1.3) can be true or false depending on their context, and equation (1.4) is obviously true. But what exactly does equation (1.1) mean? Does it mean that $5+4$ is indistinguishable from 9 and that we lose no information when we replace the symbols $5+4$ by the symbol 9? If it meant that, then equation (1.1) and (1.4) would be the same equation. But they are clearly not the same equation as $9=9$ has always been obvious to us, yet we had to spend months in primary school learning that $5+4=9$.

> **Conclusion.** After all our years in school we still don't really know what $5+4=9$ means, and we are unable to decide whether the equation $10+11=9$ is true or false.

The equations

$$10+11=9 \quad \text{on a 12-hour clock} \tag{1.5}$$

$$298+71=9 \quad \text{on a 360-degree compass} \tag{1.6}$$

are correct, as are

$$\begin{aligned} 10+11 &= 9 \quad \text{on a 12-month calendar} && (1.7)\\ 298+71 &= 9 \quad \text{on the 360-day calendar used in finance.} && (1.8) \end{aligned}$$

It is convenient to introduce a notation and vocabulary that allows us to treat equations (1.5) and (1.7) as being essentially the same equation. The notation we use is:

$$10+11 \equiv 9 \qquad \text{mod } 12 \tag{1.9}$$

The vocabulary we use is:

$$10+11 \quad \textit{is congruent to} \quad 9 \quad \textit{modulo} \quad 12 \tag{1.10}$$

We might also write

$$10+11 \not\equiv 8 \qquad \text{mod } 12 \tag{1.11}$$

and say

$$10+11 \quad \textit{is not congruent to} \quad 8 \quad \textit{modulo} \quad 12\ . \tag{1.12}$$

All of the above discussion is summarized in the following.

Definition 1.0.1 For integers a,b,c,m we write

$$a+b \equiv c \qquad \text{mod } m \tag{1.13}$$

to mean that $(a+b)-c$ is an integer multiple of m.

1.1 Addition

We are now in a position to make calculations such as the following.

$$\begin{aligned} 13+15 &\equiv 4 && \text{mod } 24 && (1.14)\\ 13+15 &\not\equiv 4 && \text{mod } 23 && (1.15)\\ 13+15 &\equiv 5 && \text{mod } 23 && (1.16)\\ 7+23 &\equiv 7 && \text{mod } 23 && (1.17)\\ 23 &\equiv 0 && \text{mod } 23 && (1.18)\\ 7+16 &\equiv 0 && \text{mod } 23 && (1.19) \end{aligned}$$

1.2 Subtraction

In light of equation (1.19) we write:

$$-7 \equiv 16 \qquad \text{mod } 23 \tag{1.20}$$

One way to understand equation (1.20) is to note that it refers to an arithmetic calculation on a 23-hour clock, and that on such a clock there are only twenty-three numbers, namely the numbers $0,1,2,\cdots,22$. So for instance, we don't allow 25 as an answer to any calculation in this context as it doesn't appear on a 23-hour clock, but we do allow 2 and note that the equation

$$25 \equiv 2 \qquad \text{mod } 23 \tag{1.21}$$

holds. So in place of 25 we write 2. Now -7 does not appear on a 23-hour clock, so what should we write in place of it? Well whatever -7 is, it should satisfy:

$$7+(-7)\equiv 0 \qquad \text{mod } 23 \tag{1.22}$$

In other words, -7 should be a number on the clock which, when added to 7, yields the answer 0. Equation (1.19) tells us that 16 is such a number. A little experimentation shows that 16 is the only such number. Hence we arrive at (1.20).

We are now in a position to make the following calculations.

$$37+(-8)\equiv 3 \qquad \text{mod } 26 \tag{1.23}$$

$$5+8\equiv -4 \qquad \text{mod } 9 \tag{1.24}$$

$$5-8\equiv 6 \qquad \text{mod } 9 \tag{1.25}$$

1.3 Multiplication

Multiplication on a clock is no problem. For instance:

$$7\times 8\equiv 6 \qquad \text{mod } 10 \tag{1.26}$$

$$7\times(-8)\equiv 4 \qquad \text{mod } 10 \tag{1.27}$$

$$5\times 13\equiv 1 \qquad \text{mod } 16 \tag{1.28}$$

1.4 Division

What should we mean by 5^{-1} on a 16-hour clock? The most reasonable interpretation is that the multiplicative inverse of 5 should be one of the numbers on the clock which, when multiplied by 5, yields 1. Equation (1.28) shows that 13 is one such number, and a bit of experimentation establishes that this is the only such number on the clock. So we write:

$$5^{-1}\equiv 13 \qquad \text{mod } 16 \tag{1.29}$$

1.5 Is there any point to all this?

Clock arithmetic is used quite a lot. As one example, suppose we wanted to order the book

> *The Famous Five - Five on a Treasure Island* by Enid Blyton

from the University library. We would simply send the books's 10-digit International Standard Book Number (ISBN) $034-002-423-2$ to the library acquisitions office. In fact, any (older) book is uniquely identified by the first nine digits of its ISBN. So we could actually just send the acquisitions office the number $034-002-423$, and we would receive the book. Well, we'd receive the book assuming that the librarian forwarded our nine digits corectly. Human's are prone to making mistakes. The librarian might get one of the nine digits wrong, say the last one, and forward $034-002-421$ which uniquely identifies the book

> *Shadowers* by Donald Hamilton

about a political maniac with a scheme to shadow and kill prominent public figures in a takeover plot and who can only be stopped by using a beautiful woman as bait. The reason for the tenth digit is that a publisher would immediately know that the 10-digit number 034-002-421-2 is not a valid ISBN number and would ask the library for correct information before sending out any book.

The last digit of a 10-digit ISBN $x_1x_2x_3\dots x_{10}$ is chosen so that the equation

$$x_1 + 2x_2 + 3x_3 + \cdots + 10x_{10} \equiv 0 \qquad \text{mod } 11 \tag{1.30}$$

holds. If this equation does not hold then the 10-digit number is not a valid ISBN and must contain an error. The final digit can be computed from the first nine digits using the formula

$$x_{10} \equiv -10^{-1}(x_1 + 2x_2 + 3x_3 + \cdots + 9x_9) \qquad \text{mod } 11 \tag{1.31}$$

on an 11-hour clock. If $x_{10} \equiv 10 \quad \text{mod } 11$ then it is represented by the symbol X in the ISBN.

We can calculate the check digit x_{10} for the *Shadowers* book as follows.

$$\begin{aligned}
x_{10} &\equiv -10^{-1}(1\times 0+2\times 3+3\times 4+4\times 0+5\times 0+6\times 2+7\times 4+8\times 2+9\times 1) && \text{mod } 11\\
&\equiv -10(6+12+12+28+16+9) && \text{mod } 11\\
&\equiv (6+1+1+6+5+9) && \text{mod } 11\\
&\equiv 28 && \text{mod } 11\\
&\equiv 6 && \text{mod } 11
\end{aligned}$$

2. Euclid’s algorithm and bank accounts

What is the best was to find the inverse k of 15 modulo 26? Otherways put, what is the best way of finding an integer k such that the equation

$$15 \times k \equiv 1 \quad \text{mod } 26 \quad \text{mod} \tag{2.1}$$

holds? One possibility is to run through the numbers $0, 1, 2, \ldots, 25$

$$\begin{array}{lll}
15 \times 0 & \equiv 0 & \text{mod } 26 \\
15 \times 1 & \equiv 1 & \text{mod } 26 \\
15 \times 2 & \equiv 4 & \text{mod } 26 \\
15 \times 3 & \equiv 19 & \text{mod } 26 \\
15 \times 4 & \equiv 8 & \text{mod } 26 \\
 & \vdots &
\end{array}$$

until we find a k satisfying equation (2.1). We’ll describe an alternative method which is better in two respects. The alternative method will be more efficient, particularly when applied to finding inverses module a number significantly larger than 26. Secondly, the alternative method provides mathematical insight into inverses in clock arithmetic.

2.1 Euclidean algorithm

The *greatest common divisor* of two integers m and n is defined to be the smallest positive integer d such that both m and n are integer multiples of d. We denote this greatest common divisor by $gcd(m,n)$. For example, $gcd(30,36) = 6$.

The alternative method begins by using a procedure known as the *Euclidean algorithm* to find the greatest common divisor of 15 and 26. Clearly we don’t really need any algorithm to determine $gcd(15,26) = 1$. But we can use the inner workings of the Euclidiean algorithm to find a solution k to (2.1). The easiest way to explain the algorithm is simply to illustrate its use.

To calculate $gcd(108,46)$ we procede as follows.

$$\mathbf{108} = 2.\mathbf{46} + \mathbf{16} \tag{2.2}$$
$$\mathbf{46} = 2.\mathbf{16} + \mathbf{14} \tag{2.3}$$
$$\mathbf{16} = 1.\mathbf{14} + \mathbf{2} \tag{2.4}$$
$$\mathbf{14} = 7.\mathbf{2} + \mathbf{0} \tag{2.5}$$

The algorithm stops when we hit a remainder of 0. The penultimate remainder, namely 2, divides all numbers on the left hand side of the equalities. (In this example we can simply observe this; but a simple inductive argument can be used to prove that the penultimate remainder will divide all numbers on the left of the equalities in every example.) In particular, the penultimate remainder divides both 108 and 46 and is thus a common divisor of the two numbers.

In order to establish that the penultimate remainder is the greatest common divisor of 108 and 46 we start with the penultimate equation (2.4) of the algorithm, and procede to use the subsequent equations in turn, as follows.

$$\mathbf{2} = \mathbf{16} - 1.\mathbf{14} \tag{2.6}$$
$$= \mathbf{16} - 1.(\mathbf{46} - 2.\mathbf{16}) = 3.\mathbf{16} - 1.\mathbf{46} \tag{2.7}$$
$$= 3.(\mathbf{108} - 2.\mathbf{46}) - 1.\mathbf{46} = -3.\mathbf{46} + 3.\mathbf{108} \tag{2.8}$$

This manipulation is summarized by the following exprssion.

$$\mathbf{2} = 3.\mathbf{108} - 3.\mathbf{46} \tag{2.9}$$

In particular, any common divisor of 46 and 108 must devide 2. We conclude that the penultimate remainder, 2, is the *greatest* common divisor.

2.2 Bézout’s identity

Note that equation (2.9) is an expression of $gcd(108,46)$ as an integer combination of 108 and 46. An expression of $gcd(m,n)$ as an integer combination $gcd(m,n) = a.m + b.n$ will always arise in this way for any integers m,n and is known as *Bézout’s Identity*. The integers a and b are not unique.

Keeping in mind our initial question of finding $15^{-1} \mod 26$, let us find a Bézout identity for $gcd(26,15)$. We begin with the Euclidean algorithm.

$$\mathbf{26} = 1.\mathbf{15} + \mathbf{11} \tag{2.10}$$
$$\mathbf{15} = 1.\mathbf{11} + \mathbf{4} \tag{2.11}$$
$$\mathbf{11} = 2.\mathbf{4} + \mathbf{3} \tag{2.12}$$
$$\mathbf{4} = 1.\mathbf{3} + \mathbf{1} \tag{2.13}$$
$$\mathbf{3} = 3.\mathbf{1} + \mathbf{0} \tag{2.14}$$

The penultimate remainder is 1 which equals $gcd(26,15)$. We now rewrite the workings of the algorithm, starting from the penultimate equation (2.13).

$$\mathbf{1} = \mathbf{4} - 1.\mathbf{3} \tag{2.15}$$
$$= \mathbf{4} - 1.(\mathbf{11} - 2.\mathbf{4}) = 3.\mathbf{4} - 1.\mathbf{11} \tag{2.16}$$
$$= 3.(\mathbf{15} - 1.\mathbf{11}) - 1.\mathbf{11} = 3.\mathbf{15} - 4.\mathbf{11} \tag{2.17}$$
$$= 3.\mathbf{15} - 4.(\mathbf{26} - 1.\mathbf{15}) = 7.\mathbf{15} - 4.\mathbf{26} \tag{2.18}$$

We thus have the Bézout identity

$$\mathbf{1} = 7.\mathbf{15} - 4.\mathbf{26} \tag{2.19}$$

2.3 Inverses in clock arithmetic

Expressing (2.19) on a 26-hour clock we find

$$7 \times 15 \equiv 1 \mod 26 \tag{2.20}$$

which yields the required inverse.

$$15^{-1} \equiv 7 \mod 26$$

This method of finding the inverse of $n \mod m$ clearly works whenever $gcd(m,n) = 1$. On the other hand, if n has some inverse k modulo m then $nk \equiv 1 \mod m$, which means there exists an integer ℓ for which the following holds.

$$nk = 1 + \ell m \tag{2.21}$$

In this case we must have $gcd(m,n) = 1$. We have just proved the following useful result.

Theorem 2.3.1 The integer n has an inverse on an m-hour clock if, and only if, $gcd(m,n) = 1$.

In particular, none of the integers $0, 2, 4, 6, 8, 10, 12, 13, 14, 16, 18, 20, 22, 24$ has an inverse modulo 26.

2.4 Bank account numbers

A bank account is identified by its International Bank Account Number (IBAN). This has the form

GB 82 WEST 123456 98765432

where *GB* is the country code, *WEST* is derived from the owner's name, *123456* is the bank sort code, and *98765432* is the account number. The digits *82* are check digits. The IBAN is converted to an integer by first rearranging it in the form

WEST 123456 98765432 GB 82

and then replacing each letter with two digits according to the scheme $A \sim 10$, $B \sim 11, \ldots, Z \sim 35$. In the given example we get the following integer.

$$n = 32142829\ 123456\ 98765432\ 1611\ 82$$

The check digits 82 are chosen to ensure

$$n \equiv 1 \mod 97.$$

The check numbers c in this example would have been chosen using the following formula.

$$c \equiv 1 - (32142829\ 123456\ 98765432\ 1611) \times 100 \equiv 82 \mod 97 \tag{2.22}$$

To use the formula we need an efficient method for calculating modulo 97. One handy method uses the observation that $100 \equiv 3 \mod 97$. Suppose for instance that we want to find the value of 31415927 modulo 97. We can procede as follows.

$$\begin{aligned}
31415927 &= 31 \times 10^6 + 41 \times 10^4 + 59 \times 10^2 + 27 && (2.23)\\
&\equiv 31 \times 3^3 + 41 \times 3^2 + 59 \times 3 + 27 \mod 97 && (2.24)\\
&\equiv 93 \times 3^2 + 26 \times 3 + 80 + 27 \mod 97 && (2.25)\\
&\equiv -4 \times 9 - 19 - 17 + 27 \mod 97 && (2.26)\\
&\equiv -45 \mod 97 && (2.27)\\
&\equiv 52\,. && (2.28)
\end{aligned}$$

3. War, finance and arithmetic

Cryptography is the practice and study of techniques for sending a message from a *sender* to a *receiver* across a *public communications channel* in such a way that if the message is intercepted by a third party as it passes through the channel then that third party will not be able to read the message in any meaningful way. For centuries the subject was mainly of interest as a military tool. However, its application domain has now broadened to include e-commerce and e-finance.

A classic example is the Enigma machine, which was the mainstream German cipher machine before and during the second world war. A military version is shown in Figure 3.1. An improved

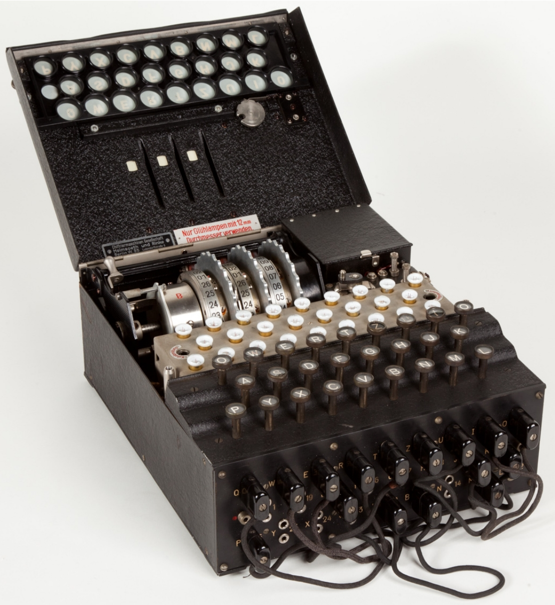


Figure 3.1: Military Enigma machine, model "Enigma I", photographed in the Museo della Scienza e della Tecnologia "Leonardo da Vinci"by Alessandro Nassiri. CC BY-SA 4.0 International

naval version was issued to U-boats. Messages could be typed into the machine at Naval Headquarters in order to produce a scrambled version for transmission to U-boats. The U-boats also carried Enigma machines which they used to unsscramble the messages. The machines could also be used to send messages from boats to Headquarters. The scrambled messages were routinely intercepted by Allied Forces who had difficulty unscrambling them. The Enigma machine had rotors (the machine of Figure 3.1 has three rotors but the improved naval version had four rotors) which could be initialized in many different positions. The initial position of the rotors influenced the scrambling. The initial rotor settings were varied on a regular basis known to both Headqurters and U-boats. This meant that if the Allies succeeded in unscambling messages on a certain day by 'breaking the code', they'd need to break the code again as soon as the initial rotor settings changed.

3.1 Basic assumptions

A sketch of the fundamental idea of cryptography is shown in Figure 3.2. We make the following assumptions.

1. Enciphering and deciphering machines are public knowledge.
2. The enciphering machine requires an *enciphering key* known only to the sender, and the deciphering machine (which is usually the same as the enciphering machine) needs a *deciphering key* known only to the receiver.
3. Enciphered messages will be intercepted.

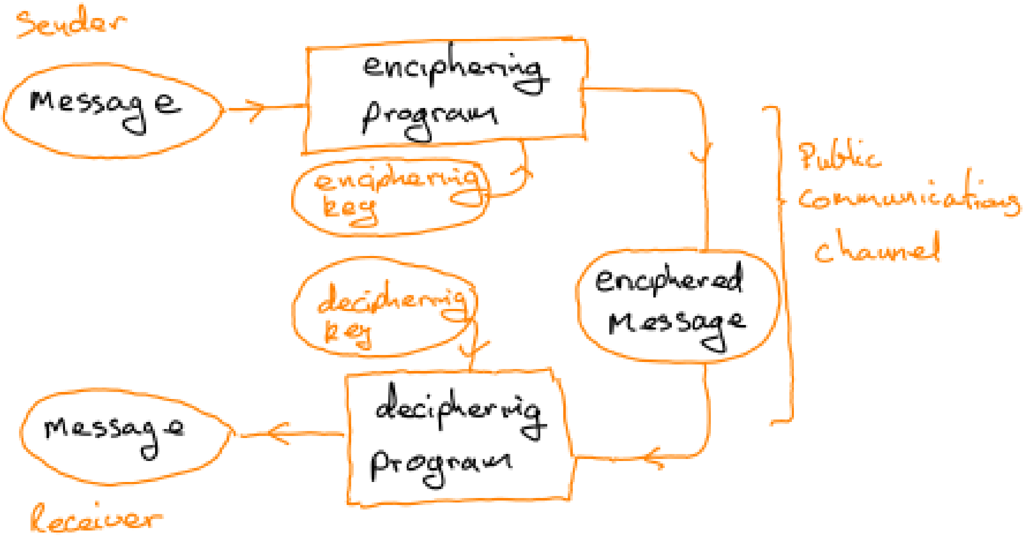


Figure 3.2: Overview of cryptography

In the case of the naval Enigma machine, assumption (1) was met in 1941 when British destroyers captured a German submarine, U-110, south of Iceland. Assumption (2) was met by the list of initial rotor positions to be used on given dates. Assumption (3) was met through the interception of German radio messeages by Allied radio operators.

There are more modern examples of cryptography. Every day, encryption is used to protect the content of web transactions, email, newsgroups, chat, web conferencing, and telephone calls as they are sent over the Internet. More will be said about the mathematics underlying this.

3.2 A first mathematical example

Suppose that a sender wants to send a secret message via email to a friend, knowing that emails are easily intercepted at various points in their journey. To keep things simple, we suppose that the message is sent using an alphabet of just 26 letters A, B, C, ..., Z, all upper case and no spaces or punctuation allowed. The message to be sent, which we refer to as the *plaintext*, is in this example the single word:

HELLO

It is convenient to index the letters by integers.

$$A \leftrightarrow 1, B \leftrightarrow 2, C \leftrightarrow 3, \ldots, Z \leftrightarrow 26$$

It is even more convenient to calculate with these integers module 26. Since

$$26 \equiv 0 \qquad \mod 26 \tag{3.1}$$

we get the correspondence $Z \leftrightarrow 0$. To denote the collection of 26 numbers on a 26-hour clock we use the symbols $\mathbb{Z}_{26}$, and refer to $\mathbb{Z}_{26}$ as the *integers modulo 26*.

For illustrative purposes we use a very simple *enciphering function* and *enciphering key*. Caveat: its simplicity means that it is extremely easy to 'break the code', and so this method should never be used in practice. We use the enciphering function

$$f_E \colon \mathbb{Z}_{26} \to \mathbb{Z}_{26}, n \mapsto an + b \tag{3.2}$$

with encipering key a suitable pair of integers $E = (a, b)$. For instance , using $E = (3, 4)$ the enciphering function becomes $f(n) = 3n + 4 \mod 26$ and our plaintext is enciphered as follows.

$$\begin{align} & H \;\; E \;\; L \;\; L \;\; O \tag{3.3}\\ \leftrightarrow & 8 \;\; 5 \;\; 12 \;\; 12 \;\; 15 \tag{3.4}\\ \to & f_E(8) \;\; f_E(5) \;\; f_E(12) \;\; f_E(12) \;\; f_E(15) \tag{3.5}\\ \to & 2 \;\; 19 \;\; 14 \;\; 14 \;\; 23 \tag{3.6}\\ \leftrightarrow & B \;\; S \;\; N \;\; N \;\; W \tag{3.7} \end{align}$$

The *ciphertext*

BSNNW

is then emailed to the friend. To decipher the message the friend uses the deciphering function

$$f_D \colon \mathbb{Z}_{26} \to \mathbb{Z}_{26}, n \mapsto \alpha n + \beta \tag{3.8}$$

with an appropriate deciphering key $D = (\alpha, \beta)$. The deciphering function and key should yield the original plaintext message. So we need the equation

$$f_D(f_E(n)) \equiv n \qquad \mod 26 \tag{3.9}$$

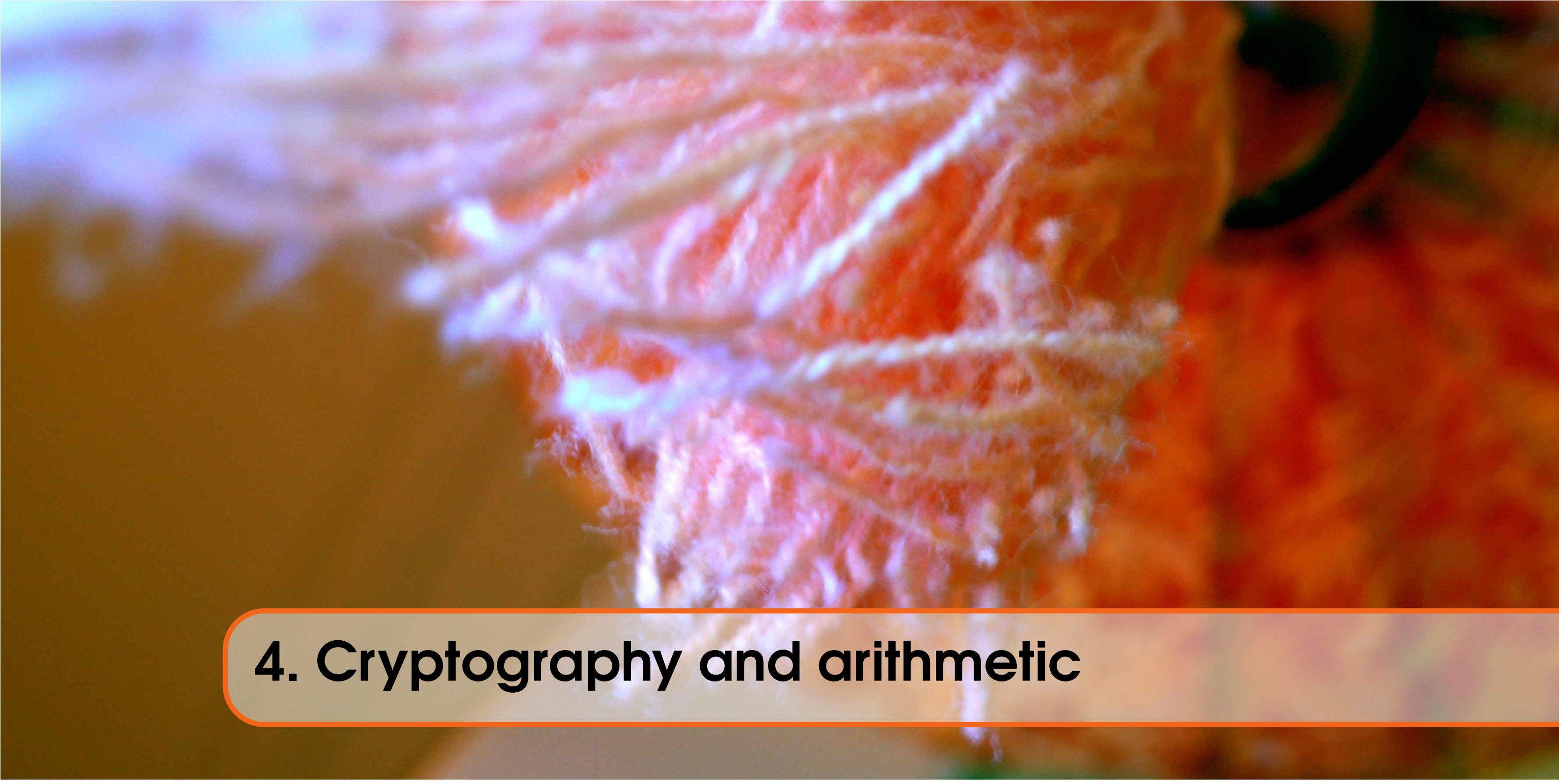
to hold. In other words, we want:

$$\begin{align} f_D(an + b) &= \alpha(an + b) + \beta \tag{3.10}\\ &= \alpha(an) + (\alpha b + \beta) \tag{3.11}\\ &\equiv n \qquad \mod 26 \tag{3.12} \end{align}$$

Thus:

$$\begin{align} \alpha &\equiv a^{-1} \qquad \mod 16 \tag{3.13}\\ \beta &\equiv -\alpha b \qquad \mod 26 \tag{3.14} \end{align}$$

For $E = (3, 4)$ the formulae (3.13) and (3.14) give $D = (9, 16)$. It is a worthwhile exercise to check that the deciphering function f_D applied with this key to the cipertext BSNNW yields the plaintext HELLO .



4. Cryptography and arithmetic

Suppose that we intercept the ciphertext

O H 7 F 8 6 B B 4 6 R 3 6 2 7 0 2 6 B B 9

and happen to know:

1. a 37-letter alphabet is used with integer correspondence:

$$0,1,\ldots,9,10 \leftrightarrow A, 11 \leftrightarrow 11,\ldots,35 \leftrightarrow Z, \_ \leftrightarrow 36$$

2. an *affine* cryptosysrem

$$f_E\colon \mathbb{Z}_{37} \to \mathbb{Z}_{37}, n \mapsto an+b$$

   is used with some unkown key $E=(a,b)$.
3. the unkown plaintext ends with the letters: 0 0 7

Let's 'break the code' by using this information to determine: i) the enciphering key; ii) and the deciphering key.

We know that $f_E(0)=B$ and $f_E(7)=9$ or, using integers in place of letters,

$$f_E(0) \equiv 11 \quad \mod 37 \tag{4.1}$$

$$0a+b \equiv 11 \quad \mod 37 \tag{4.2}$$

and

$$7a+b \equiv 9 \quad \mod 37\ . \tag{4.3}$$

From (4.2) we deduce $b \equiv 11$. Then from (4.3) we find:

$$7a \equiv -2 \quad \mod 37 \tag{4.4}$$

$$a \equiv -2 \times 7^{-1} \quad \mod 37 \tag{4.5}$$

To find 7^{-1} we can use the Euclidean algorithm

$$\begin{aligned} 37 &= 5.7+2 \quad (4.6)\\ 7 &= 3.2+1 \quad (4.7)\\ 2 &= 2.1+0 \quad (4.8) \end{aligned}$$

and Bezout's Identity

$$\begin{aligned} 1 &= 7-3.2 \quad (4.9)\\ &= 7-3(37-5.7) \quad (4.10)\\ &= 16.7-3.37 \quad (4.11) \end{aligned}$$

to deduce

$$7\times 16 \equiv 1 \qquad \text{mod } 37 \quad (4.12)$$

and finally:

$$7^{-1} \equiv 16 \qquad \text{mod } 37 \quad (4.13)$$

Now from equation (4.5) we find:

$$a \equiv -2\times 16 \equiv -16 \equiv 5 \qquad \text{mod } 37 \quad (4.14)$$

The enicphering key is thus $E=(5,11)$. Equations (3.13) and (3.14) then yield the deciphering key $D=(15,12)$.

We could use the deciphering key to decipher the cipertext. We simply need to apply the function

$$f_d(n)\colon \mathbb{Z}_{37}\to\mathbb{Z}_{37}, n\mapsto 15n+12 \quad (4.15)$$

to each letter of ciphertext. Applying f_D to the first letter of ciphertext

$$\begin{aligned} f_D(O) &\leftrightarrow f_D(24) \quad (4.16)\\ &\equiv 15\times 24+12 \qquad \text{mod } 37 \quad (4.17)\\ &\equiv 27+20 \qquad \text{mod } 37 \quad (4.18)\\ &\equiv 10 \qquad \text{mod } 37 \quad (4.19)\\ &\leftrightarrow A \quad (4.20) \end{aligned}$$

we find that the first letter of plaintext is 'A'. Subsequent letters of plaintext can be found in a similar manner.

4.1 A lesson to be learned

One lesson to learn from the above example is that when using an affine cryptosystem we should not always sign off with our name since this can help an adversary determine the enciphering key (which enables the adversary to send us 'fake news') and determine the deciphering key (from which the adversary can decipher our enrypted message).

More generally, and for the same reason, we should not send an encrypted message containing any portions whose corresponding plaintext can be identified by an adversary. This applies to many other cryptosystems too, including the Enigma machine. The German Navy had not learned this lesson! During the battle of the North Atlantic, while lying in wait for convoys of Allied ships, U-boats were collecting weather data and sending it back home. The Allies recognized these patterns and used them to decipher important German Navy messages.

If we send any large amount of ciphertext using an affine cryptosystem then there will likely be one pattern that an adversary can easily spot. If the plaintext was written in, say, English using a 26-letter alphabet then the most frequent letter in the plaintext will most likely be ‘E’ and the second most frequent letter will likely by ‘T’. The inventor of Morse code, Samuel Morse (1791–1872), needed to know frequecies of letters in English text and estimated that 12% of characters are ‘E’ and 9% are ‘T’. If the ciphertext is long enough then the most frequent letter of ciphertext will be deciphered as ‘E’ and the second most frequent letter of ciphertect will be deciphered as ‘T’. As in the above example, the adversary can solve two simultaneous equations to establish the enciphering and deciphering keys. This is one of several reasons why no affine enciphering function

$$f_E : \mathbb{Z}_N \to \mathbb{Z}_N, n \mapsto an + b \tag{4.21}$$

should ever be used in practice.

5. An old Chinese theorem

The Chinese mathematical treatize *Sunzi Suanjing*, written during the 3rd to 5th centuries AD, contains the following puzzle.

> *There are certain things whose number is unknown. Repeatedly divided by 3, the remainder is 2; by 5 the remainder is 3; and by 7 the remainder is 2. What will be the number?*

The 7th century Indian mathematician and astronomer Brahmagupta posed the following puzzle.

> *An old woman goes to market and a horse steps on her basket and crushes the eggs. The rider offers to pay for the damages and asks her how many eggs she had brought. She does not remember the exact number, but when she had taken them out two at a time, there was one egg left. The same happened when she picked them out three, four, five, and six at a time, but when she took them seven at a time they came out even. What is the smallest number of eggs she could have had?*

The first puzzle asks for 'the' integer $x \geq 0$ that simultaneously satisfies the following system of equations.

$$\left.\begin{array}{rcll} x & \equiv & 2 & \mod 3 \\ x & \equiv & 3 & \mod 5 \\ x & \equiv & 2 & \mod 2 \end{array}\right\} \tag{5.1}$$

The second puzzle asks for the smallest integer $x \geq 0$ that simultaneously satisfies the following system of equations.

$$\left.\begin{array}{rcll} x & \equiv & 1 & \mod 2 \\ x & \equiv & 1 & \mod 3 \\ x & \equiv & 1 & \mod 4 \\ x & \equiv & 1 & \mod 5 \\ x & \equiv & 1 & \mod 6 \\ x & \equiv & 0 & \mod 7 \end{array}\right\} \tag{5.2}$$

We'll provide a method for solving these two classical puzzles (and see that the first has many solutions). To avoid giving away the answers, we'll explain the method with respect to a third variant involving different numbers.

5.1 Variant of the classical puzzles

Let us find the smallest integer $x \geq 0$ that simultaneously satisfies the following system of equations.

$$\left.\begin{array}{rcll} x & \equiv & 3 & \mod 13 \\ x & \equiv & 6 & \mod 14 \\ x & \equiv & 9 & \mod 15 \end{array}\right\} \tag{5.3}$$

We construct a solution x by making a sequence of educated attempts. To start, let us set

$$a \equiv 14^{-1} \quad \mod 13 \tag{5.4}$$

$$b \equiv 15^{-1} \quad \mod 13 \tag{5.5}$$

Our first attempt at a solution to (5.3) is:

$$X = 3 \times 14 \times a \times 15 \times b \tag{5.6}$$

Observe that

$$\begin{array}{rcll} X & \equiv & 3 & \mod 13 \\ X & \equiv & 0 & \mod 14 \\ X & \equiv & 0 & \mod 15 \end{array} \tag{5.7}$$

and so our attempt satifies just the first equation.

As a second attempt set

$$c \equiv 13^{-1} \quad \mod 14 \tag{5.8}$$

$$d \equiv 15^{-1} \quad \mod 14 \tag{5.9}$$

and

$$Y = 6 \times 13 \times c \times 15 \times d\ . \tag{5.10}$$

Observe that

$$\begin{array}{rcll} Y & \equiv & 0 & \mod 13 \\ Y & \equiv & 6 & \mod 14 \\ Y & \equiv & 0 & \mod 15 \end{array} \tag{5.11}$$

and so our attempt satifies just the second equation.

As a third attempt set

$$e \equiv 13^{-1} \quad \mod 15 \tag{5.12}$$

$$f \equiv 14^{-1} \quad \mod 15 \tag{5.13}$$

and

$$Z = 9 \times 13 \times e \times 14 \times f\ . \tag{5.14}$$

Observe that

$$\begin{array}{rclr} Z & \equiv & 0 & \mod 13 \\ Z & \equiv & 0 & \mod 14 \\ Z & \equiv & 6 & \mod 15 \end{array} \tag{5.15}$$

and so our attempt satifies just the third equation.

Three failed attempts! But all is not lost. It is pretty obvious that the number

$$x = X + Y + Z \tag{5.16}$$

is a solution to the system (5.3). We can use the above formulae for X, Y and Z to find a simultaneous solution

$$x = 4410 + 15210 + 160524 = 180144 \tag{5.17}$$

to the system (5.3). But is this solution the smallest possible?

Note that $13 \times 14 \times 15 \equiv 0$ on a 13-hour clock, as well as on a 14-hour clock, and also on a 15-hour clock. So for each integer k the number

$$x = 180144 - k \times 13 \times 14 \times 15 \tag{5.18}$$

is a solution to (5.3). Moreover, any solution to (5.3) is of this form. It follows that

$$x = 180144 \quad \mod 13 \times 14 \times 15 \tag{5.19}$$

$$= 21804 \tag{5.20}$$

is the smallest positive integer satisfying (5.3).

5.2 Extracting a theorem

Why did the above method work? It worked because the inverses a, b, c, d, e, f all existed. These inverses existed because $gcd(13, 14) = 1$, $gcd(13, 15) = 1$ and $gcd(14, 14) = 1$. So from the above method we can extract the following.

Theorem 5.2.1 — Chinese Remainder Theorem. For any integers r, s, t and any integers ℓ, m, n satisfying $gcd(\ell, m) = gcd(\ell, n) = gcd(m, n) = 1$, the system of equations

$$\begin{array}{rcll} x & \equiv & r & \mod \ell \\ x & \equiv & s & \mod m \\ x & \equiv & t & \mod n \end{array}$$

has a solution.

This theorem can be generalized to systems of more than three equations.

6. Euler Phi Function and digital signatures

Two integers m and n are said to be *coprime* if $gcd(m,n) = 1$. For example, the integers 30 and 77 are coprime. The integers 6 and 21 are not coprime. A integer $p \geq 2$ is *prime* if it is coprime to all integers other than itself.

We have seen that on an m-hour clock, an integer n has an inverse n^{-1} if and only if n is coprime to m. For this reason and other reasons we are interested in the following definition.

Definition 6.0.1 *Euler's Phi function* $\phi(m)$ is defined for any positive integer m as

$$\phi(m) = \text{ the number of integers in the range } 1, 2, \ldots, m-1 \text{ that are coprime to } m \, .$$

To calculate $\phi(8)$ we could use the table

$n =$	1	2	3	4	5	6	7
$gcd(n,8) = 1$	true	false	true	false	true	false	true

to obtain $\phi(8) = 4$. To calculate $\phi(6)$ we could use the table

$n =$	1	2	3	4	5
$gcd(n,6) = 1$	true	false	false	false	true

to obtain $\phi(6) = 2$.

This tabular approach to calculating $\phi(m)$ is not so practical when n is large. For some large numbers, such as $m = 107$, the calculation of $\phi(107)$ is very straightforward: $\phi(107) = 106$. The thinking underlying this calculation generalizes to the following.

Proposition 6.0.1 If p is a prime number then $\phi(p) = p - 1$.

6.1 What is a proposition?

Mathematicians use the term *theorem* to refer to a mathematical statement that they know is true. They may be able to explain why it is true, or they may just be aware that there exists an explanation in some book or academic paper of why it is true and that this explanation has convinced many mathematicians and no credible mathematician doubts it. (An explanation of the term *credible mathematician* is beyond the scope of this book.) So a *theorem* is a mathematical statement that has a convincing explanation which meets the rigorous requirements of the mathematical community. We refer to such an explanation as a *proof*. The term *theorem* is only ever used in a mathematical context.

What then is a *proposition*? This term is used in many non-mathematical contexts. It could refer to a suggestion offered in a night club. Or a suggestion offered by a property developer to a banker. But what does the word mean when used by a mathematician?

A *proposition* is a mathematical statement that has a convincing explanation that meets the rigorous requirements of the mathematical community.

So what is the difference between a *theorem* and a *proposition*? The difference is analogous to that between a *city* and a *town*. Cities tend to be a bit larger than towns. Cities tend to be a bit higher in the unofficial hierarchy of housed communities; both cities and towns tend to be regarded as a bit more important than villages. The choice between the terms *city*, *town* and *village* is definitely country dependent. I live in a town in the West of Ireland – Oughterard – consiting of 800 inhabitants. I grew up in North Wales where we have a village – Llanfairpwllgwyngyll-gogerychwyrndrobwllllantysiliogogogoch – consisting of over 3000 inhabitants. However you measure it, the Welsh village is bigger than the Irish town.

Theorems tend to be a bit more substantial and more noteworthy than propositions. This is a subjective difference. What one mathematician might call a *theorem* another might call a *proposition*.

But what of mathematical villages? The term *lemma* is used to refer to these.

Having given a straighforward definition of the term *proposition*, let us now muddy the waters. In one branch of mathematics known as *mathematical logic* the term *proposition* just refers to any clear mathematical statement, and in this branch one allows the notion of propositions that are false as well as propositions that are true. Mathematical logic is a very specialized, though also very important, branch of mathematics. In all other branches of mathematics the above definition of *proposition* is the accepted one.

Occasionaly a proof establishing a result that has been accepted as a theorem is found to contain an error. In this case the mathematics community deals with the situation in the same way as the Catholic Church deals with broken marriage. The theorem is annuled – it is deemed never to have been a theorem.

6.2 Two more propositions

A little experimentation leads to further propositions. For instance, the easy calculations

$$\phi(2^2) = 2 \tag{6.1}$$
$$\phi(3^2) = 3 \tag{6.2}$$
$$\phi(5^2) = 20 \tag{6.3}$$
$$\phi(2^3) = 4 \tag{6.4}$$
$$\phi(2^4) = 8 \tag{6.5}$$
$$\phi(3^3) = 18 \tag{6.6}$$

might suggest a pattern. Namely, the pattern $\phi(p^k) = p^k - p^{k-1}$ when p is prime. To convert this apparent pattern into a proposition we need a convincing explanation. Here goes.

Suppose that $1 \leq n < p^k$. Then $gcd(n, p^k) > 1$ if and only if $n = pa$ with a any integer $1 \leq a \leq p^{n-1}$. There are p^{n-1} such integers a. So there are $p^n - p^{n-1}$ integers in the range $1, \ldots, p^n$ that are coprime to p^n. This establishes the following.

Proposition 6.2.1 For any prime p wnd integer $k \geq 1$ we have

$$\phi(p^k) = p^k - p^{k-1} .$$

More experimentation leads to further propositions. For instance, the calculations

$$\phi(15) = 8 \tag{6.7}$$
$$\phi(3) = 2 \tag{6.8}$$
$$\phi(5) = 4 \tag{6.9}$$

show that $\phi(3 \times 5) = \phi(3)\phi(5)$. On the other hand the calculations

$$\phi(24) = 8 \tag{6.10}$$
$$\phi(4) = 2 \tag{6.11}$$
$$\phi(6) = 2 \tag{6.12}$$

show that $\phi(4 \times 6) \neq \phi(4)\phi(6)$. The mathematical literature contains an abundance of proofs of the following proposition. It is a worthwhile excercise figuring out a proof without recourse to the literature.

Proposition 6.2.2 For any pair of coprime integers m and n the equality

$$\phi(mn) = \phi(m)\phi(n)$$

holds.

We are now in a position to quickly calculate the Euler Phi function of larger integers. For example:

$$\phi(440) = \phi(2^3 \times 3 \times 5) \tag{6.13}$$
$$= \phi(2^3)\phi(3)\phi(5) \tag{6.14}$$
$$= (2^3 - 2^2)(3-1)(5-1) \tag{6.15}$$
$$= 160 \tag{6.16}$$

6.3 Public Key Cryptography

A cryptosystem requires an enciphering key and a deciphering key. The keys need to be changed on a regular basis to maintain security. But how should sender and receiver agree on new keys? There is little point using the existing keys to encrypt and send new keys! One possibility is to pre-arrange which keys to use at given times and to record these arrangement in a codebook. The risk is that the codebook may fall into the hands of an adversary. The German Navy used this method in World War II and, sure enough, on 30 October 1942 the Allies captured the codebook. The submarine $U - 559$ was spotted at around 5am by an RAF Sunderland. Allied destroyers hunted her with depth charges for 16 hours. She was eventually damaged, and the crew had to abort the vessel. But they failed to destroy the codebbook before leaving, and it was retrieved from the sinking U-boat by three Royal Navy sailors.

In 1976 Whitfield Diffie and Martin Hellma published a paper in which they proposed the following.

Definition 6.3.1 — Diffie & Hellman. A *public key cryptosystem* is a cryptosystem with the property that someone who knows only the enciphering key can not, without a prohibitively large effort, determine how to decipher.

The idea had in fact been invented five years earlier, independently and under the name *non-secret encryption*, by James Ellis who worked for the British Govenrment Communications Headquarters (GCHQ) – an organization not so eager to publicize its work!

The affine cryptosystem described in Chapter 3 is not public key. Formulae (3.13) and (3.14) allow one to quickly compute the deciphering key from a knowledge of the enciphering key.

A cryptosystem satisfying the definition of a public key cryptosystem was invented and publicly described in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman. Rivest was a mathematics graduate from Yale, Shamir a mathematics graduate from Tel Aviv, and Adleman a mathematics graduate from Berkeley. The system has become known as *RSA cryptography* and is widely used on a daily basis by those sending messages or buying goods over the internet. It is based on properties of Euler's Phi function.

The problem of constructing a public key encryption system had in fact been given to a new recruit to GCHQ – Clifford Cocks – in 1973. Cocks had studied mathematics as an undergraduate at Cambridge and as a postgraduate at Oxford. Within a day at GCHQ he had invented what is essentially the RSA algorithm, a full four years before Rivest, Shamir and Adleman published their public key cryptosystem. This history became known only in 1997 when GCHQ declassified his invention.

6.4 Digital signatures

Before describing the RSA example of public key cryptography, we explain how any public key system

$$\begin{tikzcd} \text{plaintext} \arrow[r, bend left, "f_E"] & \text{ciphertext} \arrow[l, bend left, "f_D"] \end{tikzcd}$$

can allow a bank customer to remotely sign papers for a transaction.

The customer first needs to create an enciphering key E and deciphering key D for the cryptosystem. The customer's enciphering key E is registered with the bank, and can be made public by the customer if desired. For instance, E could be placed on a public web page. The deciphering key D is kept sercret by the customer – not even the bank is given any knowledge of D.

When the customer emails the bank asking it to send €10 000 to a supplier of goods, the bank first needs to confirm that it is indeed the customer and not some adversary who has emailed them. To confirm this, the bank generates some random message such as

Cymru_am_byth

and sends the enciphered message $f_E(\text{Cymru\_am\_byth})$ to the customer. The bank then asks the customer for details of the random message. The customer calculates

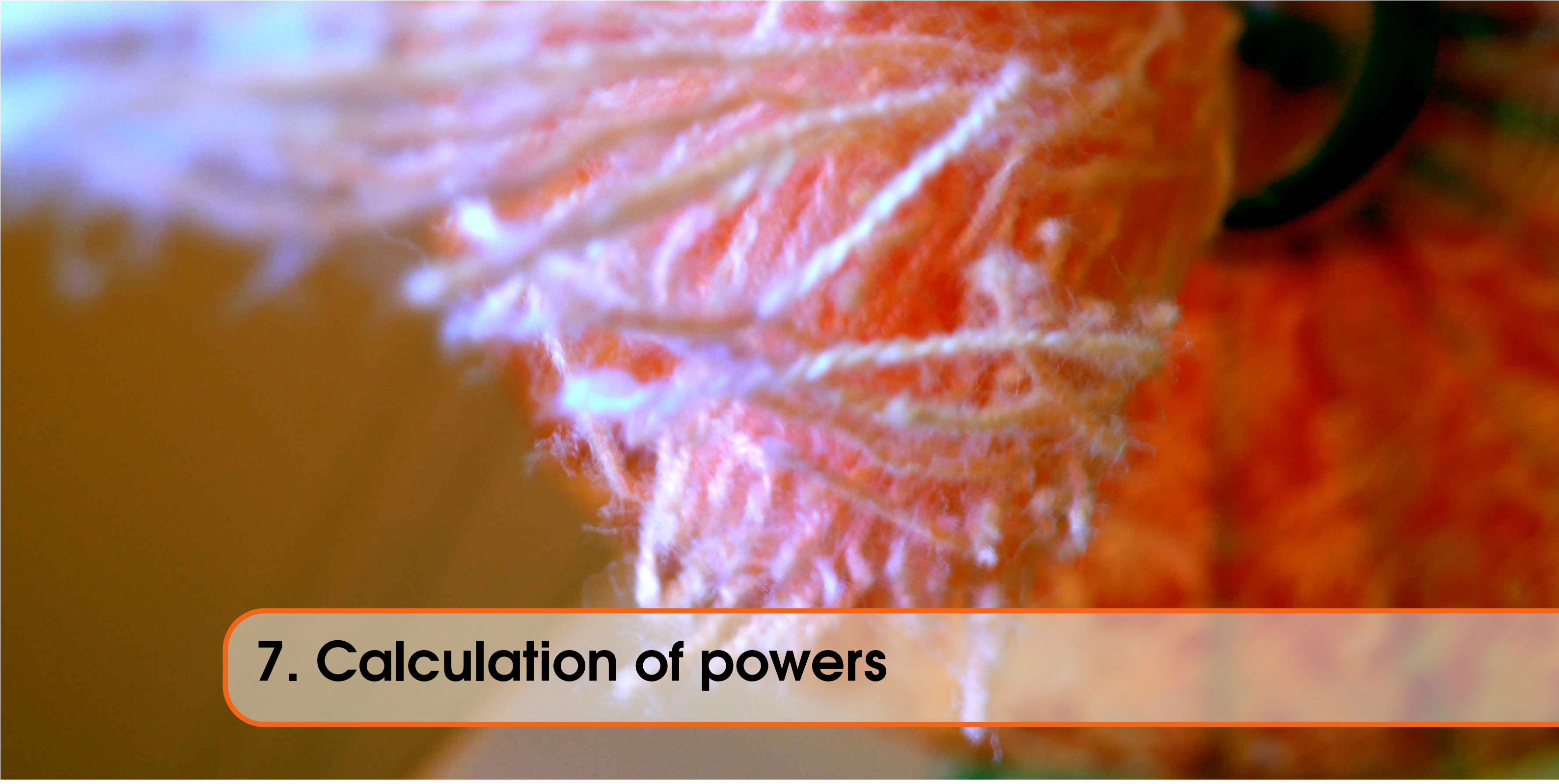
$$f_D(f_E(\text{Cymru\_am\_byth})) = \text{Cymru\_am\_byth}$$

and is able to send the random message back to the bank. Only this customer is able to do this as only the customer knows the deciphering key D.

6.5 A puzzle

Mary and Joseph have fallen in love, and Joseph wishes to send her a ring via mail. Unfortunately they live in Kleptopia where anything sent by mail will be stolen unless it is in a padlocked box.

The two of them have many padlocks, but none to which the other has a key. How can Joseph mail the ring safely to Mary?



7. Calculation of powers

To calculate 4^{30} on a 7-hour clock it would be inelegant, and inefficient, to do the following.

$$
\begin{align}
4^{30} &= 1152921504606846976 \tag{7.1}\\
&= 1 + 7 \times 164703072086692425 \tag{7.2}\\
&\equiv 1 \quad \text{mod } 7 \tag{7.3}
\end{align}
$$

There is a more elegant approach.

$$
\begin{align}
4^{30} &= ((4^2)^3)^5 \tag{7.4}\\
&\equiv (2^3)^5 \qquad \text{mod } 7 \tag{7.5}\\
&\equiv 1^5 \qquad \text{mod } 7 \tag{7.6}\\
&\equiv 1 \qquad \text{mod } 7 \tag{7.7}
\end{align}
$$

To calculate 38^{75} on a 103-hour clock we could procede as follows.

$$
\begin{align}
38^{75} &= 38^{(64+8+2+1)} \tag{7.8}\\
&= 38(38^2)(38^8)(38^{64}) \tag{7.9}\\
&\equiv 38(2)(2^4)(2^{32}) \qquad \text{mod } 103 \tag{7.10}\\
&\equiv 76(2^4)(2^8)^4 \qquad \text{mod } 103 \tag{7.11}\\
&\equiv 76(2^4)(50)^4 \qquad \text{mod } 103 \tag{7.12}\\
&\equiv 76(-3)^4 \qquad \text{mod } 103 \tag{7.13}\\
&\equiv 76 \times 81 \qquad \text{mod } 103 \tag{7.14}\\
&\equiv 79 \qquad \text{mod } 103 \tag{7.15}
\end{align}
$$

These are useful tricks for calculating powers in clock arithmetic, though not too exciting from a mathematical viewpoint.

7.1 Euler's theorem

The following result is useful and, on first encounter, probably quite surprising.

Theorem 7.1.1 — Euler's theorem. If m and a are coprime positive integers, then

$$a^{\phi(m)} \equiv 1 \qquad \text{mod } m\,.$$

To illustate the theorem for $a = 4$, $m = 9$ we calculate $\phi(9) = 6$ and note:

$$4^6 \equiv (4^2)^3 \qquad \text{mod } 9 \tag{7.16}$$
$$\equiv 7^3 \qquad \text{mod } 9 \tag{7.17}$$
$$\equiv 1 \qquad \text{mod } 9 \tag{7.18}$$

Euler's theorem is useful for calculating powers in clock arithmetic. To illustrate this let us calculate $2^{1\,000\,000}$ mod 77. First we calculate $\phi(77)$.

$$\phi(77) = \phi(7 \times 11) \tag{7.19}$$
$$= \phi(7)\phi(11) \tag{7.20}$$
$$= 6 \times 10 \tag{7.21}$$
$$= 60 \tag{7.22}$$

Next we calculate the power.

$$2^{1\,000\,000} = (2^{60})^{16666}2^{40} \tag{7.23}$$
$$\equiv 1^{16666}2^{40} \qquad \text{mod } 77 \tag{7.24}$$
$$\equiv (2^8)^5 \qquad \text{mod } 77 \tag{7.25}$$
$$\equiv 25^5 \qquad \text{mod } 77 \tag{7.26}$$
$$\equiv 9 \times 9 \times 25 \quad \text{mod } 77 \tag{7.27}$$
$$\equiv 23 \qquad \text{mod } 77 \tag{7.28}$$

We shall prove a special case of Euler's theorem and then leave the reader to try to extend this proof to the general case.

7.2 Fermat's little theorem

By taking $m = p$ a prime in Euler's theorem, and noting that $\phi(p) = p - 1$, we arrive at the following result of Pierre de Fermat.

Theorem 7.2.1 — Fermat's little theorem. For a prime p and integer a not divisible by p the equation

$$a^{p-1} \equiv 1 \qquad \text{mod } p$$

holds.

To explain why Fermat's little theorem is true, let a and p be integers satisfying the hypothesis of the theorem. Consider the numbers

$$a,\ 2a,\ 3a,\ \ldots,\ (p-1)a \qquad \text{mod } p \tag{7.29}$$

on a p-hour clock. We claim that no two numbers in this list are equal. For if two of the numbers in the list, say $i.a$ and $j.a$, were the same modulo p then:

$$i.a - j.a \equiv 0 \qquad \text{mod } p \tag{7.30}$$

This would imply:

$$(i-j)a \equiv 0 \qquad \text{mod } p \tag{7.31}$$

Thus $(i-j)a$ would be divisible by p. Since a is coprime to p this would mean that p divides (i-j). That in turn would mean

$$(i-j) \equiv 0 \qquad \text{mod } p \tag{7.32}$$

and consequently:

$$i \equiv j \qquad \text{mod } p \tag{7.33}$$

But since $1 \leq i, j < p$, we would then have $i = j$. Thus numbers in the list (7.29) are distinct from each other.

On taking the product of the numbers in the list, we find:

$$\begin{aligned}(a)(2a)(3a)\cdots((p-1)a) &= 1\times 2\times 3\times\cdots\times(p-1)a^{p-1} && (7.34)\\ &\equiv 1\times 2\times 3\times\cdots\times(p-1) \qquad \text{mod } p && (7.35)\end{aligned}$$

This implies

$$a^{p-1} \equiv 1 \qquad \text{mod } p \tag{7.36}$$

as required.

7.3 In readiness for RSA cryptography

The following application of Euler's theorem will be needed for a discussion of RSA cryptography.

Lemma 7.3.1 Let p and q be distinct prime numbers. Let e be an integer which is not divisible be either p or q, and set

$$d = e^{-1} \qquad \text{mod } (p-1)(q-1)\ ,$$

Then for any integer a that is not divisible by either p or q the equation

$$(a^e)^d \equiv a \qquad \text{mod } pq$$

holds.

The lemma is proved by noting:

$$\begin{aligned}(a^e)^d &= a^{ed} && (7.37)\\ &= a^{1+k(p-1)(q-1)} \quad \text{for some integer } k && (7.38)\\ &= a(a^{\phi(pq)})^k && (7.39)\\ &\equiv a(1)^k \qquad \text{mod } pq \qquad (\text{by Euler's theorem}) && (7.40)\\ &\equiv a \qquad \text{mod } pq && (7.41)\end{aligned}$$

8. Pretty good privacy

On September 20, 1983 the Massachusetts Institute of Technology was granted U.S. Patent 4,405,829 for a *Cryptographic communications system and method*. The patent lists Ron Rivest, Adi Shamir, and Leonard Adleman as the inventors of the method. The patent abstract states:

> The system includes a communications channel coupled to at least one terminal having an encoding device and to at least one terminal having a decoding device. A message-to-be-transferred is enciphered to ciphertext at the encoding terminal by encoding the message as a number M in a predetermined set. That number is then raised to a first predetermined power (associated with the intended receiver) and finally computed. The remainder or residue, C, is... computed when the exponentiated number is divided by the product of two predetermined prime numbers (associated with the intended receiver).

US export laws at the time prohibited export of this and similar cryptographic inventions. The patent has now expired and the export laws have been relaxed. So we are free to divulge the mathematical invention covered by the patent. In fact, since the UK based mathematician Clifford Cocks had already invented the method in 1973, it is likely that neither MIT nor the US government could every have won a court battle against anyone for using this mathematics outside of the USA.

To explain the invention let us suppose that we are working with plaintext over an N-letter alphabet, and that we have decided to split the plaintext up into k-letter message units and to split the ciphertext up into ℓ-letter message units. For existence, we might be using the finite alphabet

$$A \leftrightarrow 0,\ B \leftrightarrow 1,\ C \leftrightarrow 2,\ \ldots,\ Z \leftrightarrow 25$$

consisting of $N = 26$ letters, with a correspondence between the alphabet letters and the numbers in $\mathbb{Z}_{27}$. We might be using $k = 3$ so that plaintext such as

MEETMETONIGHTOK

would be split into a list

MEE, TME, TON, IGH, TOK

of six 3-letter message units. We might be using $\ell = 4$ so that ciphertext such as

XYABZTAA

would be split into a list

XYAB, ZTAA

of two 4-letter message units. We choose bijections:

$$\text{plaintext message units} \longleftrightarrow \text{integers } 0 \leq i < N^k \tag{8.1}$$

$$\text{ciphertext message units} \longleftrightarrow \text{integers } 0 \leq i < N^\ell \tag{8.2}$$

Convenient bijections are obtained by regarding message units as polynomials in N, as in the following example.

$$Y\ E\ S \longleftrightarrow 24.\mathbf{26}^2 + 4.\mathbf{26} + 18.\mathbf{26}^0 = 16346 \tag{8.3}$$

With thses preliminaries out of the way, we now describe the RSA public key cryptosystem.

8.1 RSA cryptosystem

- A user chooses two distinct random prime numbers p and q (each of around 1000 digits to be safe against current computer capabilities).
- The user chooses an integer e that is not divisible by either p or q, and uses the Euclidean algorithm to compute:

$$d \equiv e^{-1} \quad \mod (p-1)(q-1) \tag{8.4}$$

- The user computes the product:

$$n = pq \tag{8.5}$$

- The user publishes the enciphering key $E = (n, e)$ and keeps secret the second component of the deciphering key $D = (n, d)$.
- A plaintext message unit, corresponding to an integer a, is enciphered as:

$$f_E(a) \equiv a^e \qquad \mod n \tag{8.6}$$

- A ciphertext message unit, corresponding to an integer a, is deciphered as:

$$f_D(a) \equiv a^d \qquad \mod n \tag{8.7}$$

Lemma (7.3.1) ensures that $f_D(f_E(a)) = a$. That is, the ciphertext gets deciphered into the original plaintext.

To illustrate the cryptosystem let us continue with the above alphabet of $N = 26$ letters, plaintext message units of length $k = 3$ and ciphertext message units of length $\ell = 4$.cryptosystem let us continue with the above alphabet of $N = 26$ letters, plaintext message units of length $k = 3$ and ciphertext message units of length $\ell = 4$.

Suppose that Bob wants to send Alice the message YES . He looks up Alice's public key, which we take to be

$$E_{Alice} = (N, e) \tag{8.8}$$

$$= (46927, 39423) \tag{8.9}$$

where the value of n is kept unrealistically small for illustrative purposes. Bob computes

$$\begin{aligned} f_{E_{Alice}}(\text{YES}) &\leftrightarrow f_{E_{Alice}}(16346) && (8.10)\\ &= 16346^{39423} \mod 46927 && (8.11)\\ &= 21166 && (8.12)\\ &= 1.\mathbf{26}^3 + 5.\mathbf{26}^2 + 8.\mathbf{26} + 2.\mathbf{26}^0 && (8.13)\\ &\leftrightarrow \text{BFIC} && (8.14) \end{aligned}$$

Bob thus sends the ciphertext BFIC to Alice.

On receiving thr ciphertext Alice deciphers it using her secret deciphering key.

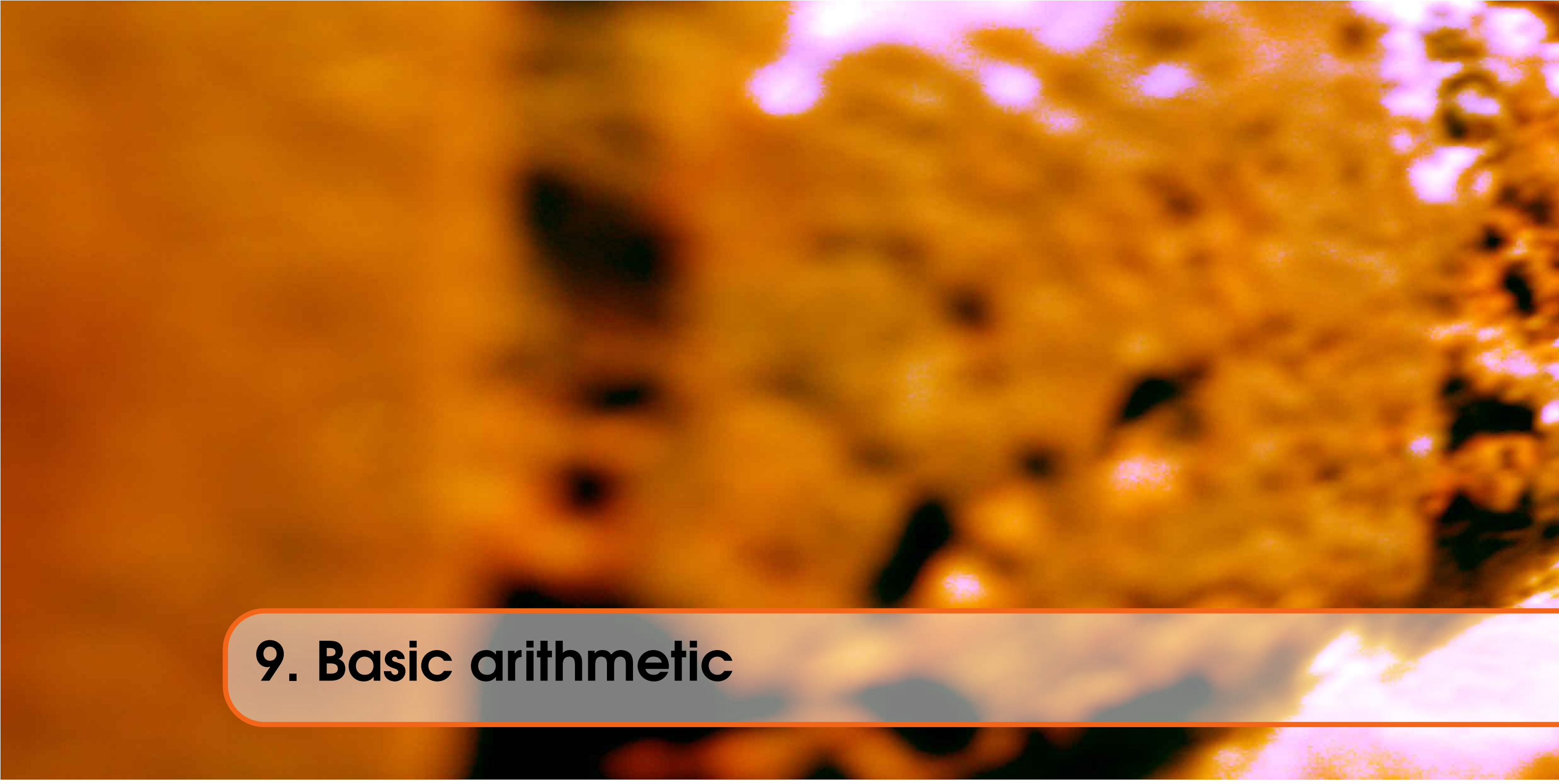
8.2 Two remarks

1. Currently the only known method for the task of calculating $d \equiv e^{-1} \mod \phi(n)$ involves factoring $n = pq$ to determine the primes p and q that Alice chose when constructing her keys.
2. Currently the task of factoring n as a product of primes is prohibitively time consuming when p and q are well-chosen large primes.

The search for new mathematics relating to these two tasks is an active area of current research.

II

Matrices



9. Basic arithmetic

Clock arithmetic is very much like the arithmetic taught in school. In particular, the following rules

$$\begin{array}{lcll}
a+b & = & b+a & \text{(commutative addition)}\\
ab & = & ba & \text{(commutative multiplication)}\\
(a+b)+c & = & a+(b+c) & \text{(associative addition)}\\
(ab)c & = & a(bc) & \text{(associative multiplication)}\\
a(b+c) & = & ab+ac & \text{(distributivity)}
\end{array}$$

hold for any $a,b,c \in \mathbb{Z}_m$ and integer $m \geq 1$. We'll now study an arithmetic where multiplication is not necessarily commutativity.

9.1 Matrices

A *matrix* is an array of numbers aranged neatly in rows and columns. The rows all have the same length. The columns all have the same length. Some examples are

$$\begin{pmatrix} 1 & 2 & 5 \\ -2 & 3 & 10 \end{pmatrix} \qquad 2 \times 3 \text{ matrix}, \tag{9.1}$$

$$\begin{pmatrix} \frac{1}{2} & -\sqrt{2} \\ 3 & 7 \end{pmatrix} \qquad 2 \times 2 \text{ matrix}, \tag{9.2}$$

$$\begin{pmatrix} 1 & 2 & 3 & -4 & 5 \end{pmatrix} \qquad 1 \times 5 \text{ matrix, also called a } \textit{row vector}, \tag{9.3}$$

$$\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \qquad 3 \times 1 \text{ matrix, also called a } \textit{column vector.} \tag{9.4}$$

9.2 Matrix addition

Two $m \times n$ matrices A, B are *added* by adding corresponding entries.

■ **Example 9.1**

$$\begin{array}{ccccc}
A & & B & = & A+B \\
\begin{pmatrix} 17 & 22 & 42 \\ 6 & 18 & 5 \end{pmatrix} & + & \begin{pmatrix} 1 & 2 & 3 \\ -6 & 1 & 5 \end{pmatrix} & = & \begin{pmatrix} 18 & 24 & 45 \\ 0 & 19 & 10 \end{pmatrix} \\
\begin{pmatrix} -2 & 3 \\ 1 & 4 \end{pmatrix} & + & \begin{pmatrix} 7 & 6 \\ 5 & 4 \end{pmatrix} & = & \begin{pmatrix} 5 & 9 \\ 6 & 8 \end{pmatrix} \\
\begin{pmatrix} -2 & 3 & 1 & 0 \end{pmatrix} & + & \begin{pmatrix} 7 & 6 & -1 & -2 \end{pmatrix} & = & \begin{pmatrix} 5 & 9 & 0 & -2 \end{pmatrix} \\
\begin{pmatrix} 17 & 22 & 42 \\ 6 & 18 & 5 \end{pmatrix} & + & \begin{pmatrix} 7 & 6 \\ 5 & 4 \end{pmatrix} & & \text{not defined}
\end{array}$$

■

Given a matrix A we write $-A$ to denote the matrix obtained by placing a minux in front of each entry of A.

■ **Example 9.2**

$$\begin{array}{cc}
A & -A \\
\begin{pmatrix} 2 & -3 \\ 4 & 7 \end{pmatrix} & \begin{pmatrix} -2 & 3 \\ -4 & -7 \end{pmatrix}
\end{array}$$

■

For any $m \times n$ matrix A the sum $A + (-A)$ yields the $m \times n$ matrix whose entries are all zero. Such a matrix is called a *null matrix* or *zero matrix*. We write

$$A + (-A) = 0 \tag{9.5}$$

where here 0 denotes a null matrix.

9.3 Multiplication of a column vector by row vector

Let

$$R = (a_1, a_2, \dots, a_n) \tag{9.6}$$

be a row vector of length n. Let

$$C = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{9.7}$$

be a column vector of length n. The *product* $R.C$ is a number defined as:

$$R.C = (a_1, a_2, \dots, a_n) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + a_2 b_2 + \cdots + b_n c_n \tag{9.8}$$

■ **Example 9.3** For example:

$$R.C = (-2, 3, 7) \begin{pmatrix} 8 \\ 5 \\ 9 \end{pmatrix} = -2 \times 8 + 3 \times 5 + 7 \times 9 = 62 \tag{9.9}$$

■

If we regard the number $R.C$ as a 1×1 matrix then we obtain the 'formula':

$$(1 \times n \text{ matrix}) \times (n \times 1 \text{ matrix}) = (1 \times 1 \text{ matrix}) \tag{9.10}$$

We'll see below that it is also possible to define the product $C.R$, though in this case our definition will yield an $n \times n$ matrix. In particular, we'll get $R.C \neq C.R$ whenever $n > 1$.

9.4 Matrix multiplication

Let A be an $m \times n$ matrix. We can view A as a list of m row vectors. We write R_i^A to denote the ith row vector of A.

Let B be an $n \times p$ matrix. We can view B as a list of p column vectors. We write C_j^A to denote the jth column vector of B.

For each $1 \leq i \leq m$ and $1 \leq j \leq n$ we can multiply the $i$the row of A and jth column of B to obtain the number $R_i^A C_j^B$. These numbers are used define the matrix product AB as:

$$AB = \begin{pmatrix} R_1^A C_1^B & R_1^A C_2^B & R_1^A C_3^B & \cdots & R_1^A C_p^B \\ R_2^A C_1^B & R_2^A C_2^B & R_2^A C_3^B & \cdots & R_2^A C_p^B \\ R_3^A C_1^B & R_3^A C_2^B & R_3^A C_3^B & \cdots & R_3^A C_p^B \\ \vdots & & & & \vdots \\ R_m^A C_1^B & R_m^A C_2^B & R_m^A C_3^B & \cdots & R_m^A C_p^B \end{pmatrix} \tag{9.11}$$

Thus AB is an $m \times p$ matrix whose entry in the ith row and jth column is the product of the ith row of A and jth column of B. We have the 'formula':

$$(m \times n \text{ matrix}) \times (n \times p \text{ matrix}) = (m \times p \text{ matrix}) \tag{9.12}$$

■ **Example 9.4** As an example of this definition of matrix multiplication we have:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} -1 & 2 & 1 \\ 1 & 3 & 3 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 14 & 7 \\ 1 & 35 & 19 \end{pmatrix} \tag{9.13}$$

■

9.5 Algebraic properties

The equalities

$$\begin{array}{lcll} A+B & = & B+A & \text{(commutative addition)} \\ (A+B)+C & = & A+(B+C) & \text{(associative addition)} \end{array}$$

hold for any three $m \times n$ matrices A, B, C.

The equalities

$$\begin{array}{lcll} (AB)C & = & A(BC) & \text{(associative multiplication)} \\ A(B+B') & = & AB+AB' & \text{(distributivity)} \end{array}$$

hold for any $m \times n$ matrix A, $n \times p$ matrices B, B', and $p \times q$ matrix C.

10. Scalars, division, affine cryptography

For a number k and $m \times n$ matrix A we write

$$kA \tag{10.1}$$

to denote the $m \times n$ matrix obtained from A by muktiplying each of its entries by k.

■ **Example 10.1**

$$-2\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} -2 & -4 & -6 \\ -8 & -10 & -12 \end{pmatrix} \tag{10.2}$$

■

In this setting we refer to the number k as a *scalar*, and we refer to (10.1) as *scalar multiplication*.

10.1 Algebraic properties of scalar multiplication

The equalities

$$\begin{array}{rcll} k(A+B) & = & (kA)+(kB) & \text{(distributivity)} \\ (k+\ell)A & = & kA+\ell A & \text{(distributivity)} \end{array}$$

hold for any numbers k, ℓ and any $m \times n$ matrices A, B. The equalities

$$\begin{array}{rcll} k(AB) & = & (kA)B & \text{(associativity)} \\ (k\ell)A & = & k(\ell A) & \text{(associativity)} \\ k(AB) & = & A(kB) & \text{(scalar commutativity)} \\ k(\ell A) & = & \ell(kA) & \text{(scalar commutativity)} \end{array}$$

hold for any numbers k, ℓ and any $m \times n$ matrices A, B.

10.2 Identity matrices

Let A be an $n \times n$ matrix. Such a matrix is said to be *square*. A *diagonal entry* of A is an entry lying in the ith row and ith column for some $1 \leq i \leq n$. An entry of A is said to be *non diagonal* if it is not a diagonal entry. An *identity matrix* is a square matrix whose diagonal entries are all equal to 1, and whose non diagonal entries are all equal to 0. We denote such a matrix by I_n, or just by I if the value of n is clear from the context.

■ **Example 10.2**

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{10.3}$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10.4}$$

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{10.5}$$

■

The equalities

$$IA = A = AI \tag{10.6}$$

hold for any $n \times n$ matrix A and $I = I_n$.

10.3 Inverse matrices

The following definition repeats the idea we used in clock arithmetic for introducing a notion of dibision into our new setting.

Definition 10.3.1 Let A be a square $n \times n$ matrix. An $n \times n$ matrix B is said to be an *inverse* to A if the equalitiies

$$AB = I = BA$$

hold.

■ **Example 10.3** The matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 5 \\ 3 & 8 & 6 \end{pmatrix}, \qquad B = \begin{pmatrix} 10 & -12 & 5 \\ -3 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix}$$

satisfy

$$AB = I = BA \tag{10.7}$$

and so B is an inverse to A. Is there any other inverse to A? Well, if C were also an inverse to A we'd have:

$$C(AB) = C(I) \tag{10.8}$$

$$(CA)B = C \tag{10.9}$$

$$IB = C \tag{10.10}$$

$$B = C \tag{10.11}$$

■

This little calculation establishes the following fundamental result.

Theorem 10.3.1 A square matrix A has at most one inverse.

If a matrix A has an inverse then we denote this inverse by A^{-1}. Clearly some matrices, such as the zero matrix, have no inverse since we can never have $A0 = I$.

For the moment let us focus on 2×2 matrices. For an arbitrary 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{10.12}$$

we have:

$$\begin{aligned} A\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} && (10.13) \\ &= \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} && (10.14) \\ &= (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} && (10.15) \\ &= (ad-bc)I && (10.16) \end{aligned}$$

Similarly, we have:

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} A = (ad-bc)I \tag{10.17}$$

Equations (10.16) and (10.17) establish the following formula for calculating inverses of 2×2 matrices.

Theorem 10.3.2 A matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has an inverse if and only if the number $ad - bc$ has an inverse. If this number has an inverse then

$$A^{-1} = (ad-bc)^{-1}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

■ **Example 10.4** The matrix

$$A = \begin{pmatrix} 1 & 4 \\ 5 & 7 \end{pmatrix} \tag{10.18}$$

has inverse

$$\begin{aligned} A^{-1} &= (1 \times 7 - 5 \times 4)^{-1}\begin{pmatrix} 7 & -4 \\ -5 & 1 \end{pmatrix} && (10.19) \\ &= \begin{pmatrix} \frac{-7}{13} & \frac{4}{13} \\ \frac{5}{13} & \frac{-1}{13} \end{pmatrix}. && (10.20) \end{aligned}$$

■

10.4 Affine cryptography

The cryptosystem $f_E : \mathbb{Z}_N \to \mathbb{Z}_N, n \mapsto an + b$ introduced in Section 3.2 has the merit of being simple to implement, and efficient for enciphering long messages. But it has two serious drawbacks:

1. Frequency analysis of the letters in a long ciphertext would lead to a pair of linear simultaneous equations modulo N from which the enciphering and deciphering keys could likely be calculated very quickly. Here N denotes the length of the alphabet used.
2. The number of possibly enciphering keys $E = (a, b)$ is equal to $\phi(N)N$ where the Euler Phi function $\phi(N)$ determines the number of invertible numbers a on an N-hour clock. We refer to $\phi(N)N$ as the size of the *key space*. In our examples we used an alphabet with $N = 26$ letters, for which the key space has size:

$$\begin{aligned} \phi(26)26 &= \phi(2)\phi(13)26 && (10.21)\\ &= 12 \times 26 && (10.22)\\ &= 312 && (10.23) \end{aligned}$$

For each of the 312 possible enciphering keys one could quickly compute the corresponding deciphering key using the Euclidean algorithm with formulae (3.13) and (3.14); all deciphering keys could be applied to any ciphertext to see which one yields meanigful plaintext. In more realistic examples there would be maybe one letter for each character on a computer key board. A standard Qwerty keyboard has around $N = 104$ characters. In this case the key space would have size:

$$\begin{aligned} \phi(104)104 &= \phi(13)\phi(2^3)104 && (10.24)\\ &= 12(2^3 - 2^2)104 && (10.25)\\ &= 4992 && (10.26) \end{aligned}$$

Again, we could quickly run through all 4992 possible deciphering keys until we found one that produces meaningful plaintext.

In contrast, the RSA cryptosystem with suitably chosen enciphering and deciphering keys is a secure method for sending messages: frequecy analysis is not applicable since it is public key; it has an infinite number of possible keys. But it has the drawback that it takes a computer quite some time to encipher long messages. In practice, the RSA system is used only for sending short messages.

The simple cryptosystem of Section 3.2 is easily strengthened using matrix algebra. Instead of viewing a message as a sequence of letters, we can fix an integer $k \geq 1$ and regard plaintex as a sequence of k-tuples of letters.

For instance, when $k = 2$ we consider the plaintext

$$\mathrm{HELLO\_WORLD} \qquad (10.27)$$

as a sequence of 2-tuples:

$$\begin{pmatrix} \mathrm{H} \\ \mathrm{E} \end{pmatrix} \begin{pmatrix} \mathrm{L} \\ \mathrm{L} \end{pmatrix} \begin{pmatrix} \mathrm{O} \\ \_ \end{pmatrix} \begin{pmatrix} \mathrm{W} \\ \mathrm{O} \end{pmatrix} \begin{pmatrix} \mathrm{R} \\ \mathrm{L} \end{pmatrix} \begin{pmatrix} \mathrm{D} \\ \_ \end{pmatrix} \qquad (10.28)$$

Using the correspondence A $\leftrightarrow 0, \ldots,$ Z $\leftrightarrow 25,$ _ $\leftrightarrow 26$ we can represent this sequence as a sequence of 2×1 matrices, or column vectors:

$$\begin{pmatrix} 7 \\ 4 \end{pmatrix} \begin{pmatrix} 11 \\ 11 \end{pmatrix} \begin{pmatrix} 14 \\ 26 \end{pmatrix} \begin{pmatrix} 22 \\ 14 \end{pmatrix} \begin{pmatrix} 17 \\ 11 \end{pmatrix} \begin{pmatrix} 3 \\ 26 \end{pmatrix} \qquad (10.29)$$

And now for the exciting part! We can view (10.29) as a sequence of column vectors whose entries are from the set $\mathbb{Z}_{27}$ of integers modulo 27.

In the above discussion of matrix arithmetic we defined a matrix as an array of 'numbers'. Nowhere in the discussion did we insist that these be numbers from the set of reals $\mathbb{R}$. Everything that we have said about matrices so far holds when the entries are taken from the integers $\mathbb{Z}_N$ modulo any fixed integer N. If we have broken the plaintext into message units consisting of k-tuples of letters, then we can encipher each k-tuple using a function

$$f_E\colon (\mathbb{Z}_N)^k \to (\mathbb{Z}_N)^k, v \mapsto Av+B \qquad \bmod N \tag{10.30}$$

where the enciphering key $E=(A,B)$ consists of an invertible $k\times k$ matrix A with entries taken from $\mathbb{Z}_N$, and a $k\times 1$ column vector B again with entries taken from $\mathbb{Z}_N$. The variable v denotes a $k\times 1$ column vector of numbers from $\mathbb{Z}_N$ corresponding to a k-tuple of letters.

Our proof of Theorem 10.3.2 continues to hold when we are working modulo N. So the requirement that the matrix

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \bmod N \tag{10.31}$$

be invertible just means that the number $ad-bc$ must be invertible modulo N. By Theorem 2.3.1 this just means that we need the integer $ad-bc$ to be coprime to the integer N.

The deciphering function is

$$f_D\colon (\mathbb{Z}_N)^k \to (\mathbb{Z}_N)^k, v \mapsto \alpha v+\beta \qquad \bmod N \tag{10.32}$$

where the enciphering key $D=(\alpha,\beta)$ consists of a $k\times k$ matrix α and a $k\times 1$ column vector β. The values of α and β are determined by noting that the algebraic derivation of (3.13) and (3.14) carries over from clock arithmetic to matrix arithmetic to yield:

$$\alpha = A^{-1} \tag{10.33}$$

$$\beta = -\alpha B \tag{10.34}$$

What is the advantage to working with k-letter message units? One advantage is that for $k\geq 3$ there is no clear contender for the most frequent k-tuple of letters in English texts or texts in other languages. So frequency analysis can't be used to determine the enciphering or deciphering key. A second advantage is that for large k it has an extremely large key space. The cryptosystem (10.30) is known as an *affine cryptosystem*. It is a secure system when $k\geq 3$. It is also easy to implement on a computer and is quick at enciphering long pieces of plaintext. The question of how to send new deciphering and enciphering keys to users is easily answered these days: use a public key cryptosystem, such as the RSA system, to exchange keys. The cryptosystem considered in Section 3.2 is just an affine system with $k=1$.

The idea of using matrices to construct cryptosystems goes back to Lester S. Hill who, in 1929 invented what is now known as the *Hill cipher*. In short, the Hill cipher is an affine cryptosystem with $k=3$ and with the vector B in the enciphering key always taken to be the zero vector. When $B=0$ we say that the cryptosystem (10.30) is a *linear cryptosystem*. With modern computing power, linear systems are now considered to be insecure.

10.5 A puzzle

What is the size of the key space for the affine cryptosystem (10.30)? This is a hard question to ask in an introductory text! Is there some restriction that can be placed on N to make the question more tractable without making it so specific that it loses its mathematical interest?

11. A cyber attack

Let us use the 2-dimensional affine enciphering function

$$f_E\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \text{mod } 26 \tag{11.1}$$

to encipher the plaintext

N O A N S W E R

over the 26-letter alphabet with A ↔ 0, B ↔ 1, ..., Z ↔ 25. We begin by converting the plaintext to a sequence of column vectors over $\mathbb{Z}_{26}$, each vector of length 2.

$$\begin{pmatrix} \text{N} \\ \text{O} \end{pmatrix} \begin{pmatrix} \text{A} \\ \text{N} \end{pmatrix} \begin{pmatrix} \text{S} \\ \text{W} \end{pmatrix} \begin{pmatrix} \text{E} \\ \text{R} \end{pmatrix} \leftrightarrow \begin{pmatrix} 13 \\ 14 \end{pmatrix} \begin{pmatrix} 0 \\ 13 \end{pmatrix} \begin{pmatrix} 18 \\ 22 \end{pmatrix} \begin{pmatrix} 4 \\ 17 \end{pmatrix} \tag{11.2}$$

We then apply f_E to each column vector in turn. For the first vector we get:

$$\begin{aligned} f_E \begin{pmatrix} 13 \\ 14 \end{pmatrix} &= \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 13 \\ 14 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \text{mod } 26 && (11.3) \\ &\equiv \begin{pmatrix} 69 \\ 205 \end{pmatrix} \qquad \text{mod } 26 && (11.4) \\ &\equiv \begin{pmatrix} 17 \\ 23 \end{pmatrix} \qquad \text{mod } 26 && (11.5) \\ &\leftrightarrow \begin{pmatrix} \text{Q} \\ \text{W} \end{pmatrix} && (11.6) \end{aligned}$$

For the subsequent three vectors we get:

$$f_E\begin{pmatrix} 0 \\ 13 \end{pmatrix} = \begin{pmatrix} 14 \\ 2 \end{pmatrix} \leftrightarrow \begin{pmatrix} \text{O} \\ \text{C} \end{pmatrix} \tag{11.7}$$

$$f_E\begin{pmatrix} 18 \\ 22 \end{pmatrix} = \begin{pmatrix} 25 \\ 18 \end{pmatrix} \leftrightarrow \begin{pmatrix} \text{Z} \\ \text{S} \end{pmatrix} \tag{11.8}$$

$$f_E\begin{pmatrix} 4 \\ 17 \end{pmatrix} = \begin{pmatrix} 8 \\ 10 \end{pmatrix} \leftrightarrow \begin{pmatrix} \text{I} \\ \text{K} \end{pmatrix} \tag{11.9}$$

The ciphertext to be communicated to the receiver is thus:

Q W O C Z S I K

The deciphering function corresponding to (11.1) is:

$$f_D\colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix}^{-1} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) \qquad \text{mod } 26 \tag{11.10}$$

$$= \begin{pmatrix} 14 & 11 \\ 17 & 18 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 16 \\ 15 \end{pmatrix} \tag{11.11}$$

This 2-dimensional cryptosystem over the 26-letter alphabet has a key space of size 98804160. Assuming it takes 1 second to apply f_D, for a given choice of deciphering key D, to a ciphertext to decide if the corresponding 'plaintext' is meaningful, then it would take over three years of continuous computing to run through all possible keys. If a user felt this was insufficiently secure, then a higher dimensional affine system and/or a larger alphabet could be used. For instance, an affine cryptosystem of dimension $k = 3$ over a 103-letter alphabet has a key space of size 1411784169290494225596864. Assuming again 1 second to apply f_D and evaluate the output, it would take 4×10^{16} years of continuous computing to run through all possible keys. To place this number in some context, we note that the Earth is estimated to be only 4.5×10^9 years old.

11.1 A secure cryptosystem in inexperienced hands

Algebra is the basis of secure cryptography. But algebra can also be used to exploit the inexperience of users. To illustrate the latter, suppose that we have intercepted the ciphertext

NHVGR!_ECTFMXSST_XFPOVMJB?ZSKRTCZ_GKJDDLGKQAMCXIROMTHOOTO
VHOVIDAPY_E_XBOFXKRDPXISI?YMTAAJZDPLW

which we know has been created using an affine cryptosystem

$$f_E : v \mapsto \mathbf{A}v + \mathbf{B} \qquad \text{mod } 29 \tag{11.12}$$

of dimension $k = 2$ over the alphabet

ABCDEFGHIJKLMNOPQRSTUVWXYZ_?!

of 29 letters. We wrtie matrices in bold font in this section, so as to distinguish them from letters of the alphabet. Without any additional information about the ciphertext it would be quite difficult to determine the corresponding plaintext. However, suppose that we know the ciphertext was sent from Active Agent Karla, and that this inexperienced agent always signs off messages with:

AA_KARLA

This extra information provides tells us:

$$f_E\begin{pmatrix} \text{A} \\ \text{A} \end{pmatrix} = \begin{pmatrix} \text{A} \\ \text{A} \end{pmatrix},\ f_E\begin{pmatrix} 0 \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \bmod 29 \tag{11.13}$$

$$f_E\begin{pmatrix} \text{\_} \\ \text{K} \end{pmatrix} = \begin{pmatrix} \text{J} \\ \text{Z} \end{pmatrix},\ f_E\begin{pmatrix} 26 \\ 10 \end{pmatrix} \equiv \begin{pmatrix} 9 \\ 25 \end{pmatrix} \quad \bmod 29 \tag{11.14}$$

$$f_E\begin{pmatrix} \text{A} \\ \text{R} \end{pmatrix} = \begin{pmatrix} \text{D} \\ \text{P} \end{pmatrix},\ f_E\begin{pmatrix} 0 \\ 17 \end{pmatrix} \equiv \begin{pmatrix} 3 \\ 15 \end{pmatrix} \quad \bmod 29 \tag{11.15}$$

$$f_E\begin{pmatrix} \text{L} \\ \text{A} \end{pmatrix} = \begin{pmatrix} \text{L} \\ \text{W} \end{pmatrix},\ f_E\begin{pmatrix} 11 \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 11 \\ 22 \end{pmatrix} \quad \bmod 29 \tag{11.16}$$

Equation (11.13) tells us that

$$\mathbf{A0} + \mathbf{B} = \mathbf{0} \qquad \bmod 29 \tag{11.17}$$

and thus that the active agent has chosen $\mathbf{B}$ to be the zero vector $\mathbf{B} \equiv \mathbf{0}$. The deciphering function is thus:

$$f_D\colon v \mapsto \mathbf{A}^{-1}v \qquad \bmod 29 \tag{11.18}$$

Equations(11.15) and (11.16) can be combined into the following single matrix equation.

$$\mathbf{A}\begin{pmatrix} 0 & 11 \\ 17 & 0 \end{pmatrix} \equiv \begin{pmatrix} 3 & 11 \\ 15 & 22 \end{pmatrix} \qquad \bmod 29 \tag{11.19}$$

Multiplying both sides of equation (11.19) on the left by $\mathbf{A}^{-1}$, and multiplying both sides on the right by

$$\begin{pmatrix} 3 & 11 \\ 15 & 22 \end{pmatrix}^{-1} \qquad \bmod 29 \tag{11.20}$$

yields the equation:

$$\begin{pmatrix} 0 & 11 \\ 17 & 0 \end{pmatrix}\begin{pmatrix} 3 & 11 \\ 15 & 22 \end{pmatrix}^{-1} \equiv \mathbf{A}^{-1} \qquad \bmod 29 \tag{11.21}$$

Since matrix multiplication is not commutative it is important to distinguish between multiplying on the right and multiplying on the left in these calculations.

To evaluate the inverse matrix (11.20) we note that:

$$5 \times 22 - 15 \times 11 = -99 \equiv 17 \qquad \bmod 29 \tag{11.22}$$

We can use the Bézout identity to establish:

$$17^{-1} \equiv 12 \qquad \bmod 29 \tag{11.23}$$

Equation (11.21) now yields

$$\mathbf{A}^{-1} = \begin{pmatrix} 0 & 11 \\ 17 & 0 \end{pmatrix} 12 \begin{pmatrix} 22 & -11 \\ -15 & 5 \end{pmatrix} \qquad \bmod 29 \tag{11.24}$$

$$= \begin{pmatrix} 21 & 19 \\ 22 & 18 \end{pmatrix} \qquad \bmod 29 \tag{11.25}$$

The reader should now use the deciphering function (11.18) to determine (at least a portion of) the plaintext. This worthwhile exercise will provide plenty of practice at modular arithmetic.

12. Linear transformations of the plane

Who invented the rules for matrix multiplication? Why were they invented? The second of these questions is the easier, and will be addressed next lecture after some necessary background has been covered. The mathematics literature shows that several people *discovered* the rules involved in matrix multiplication – they arise naturally as the answer to various fundamental problems about areas, volumes, and geometric transformations. The English mathematician James Joseph Sylvester is credited with having coined the term *matrix* (which is the Latin for *womb*) in a paper of 1848. The English mathematician Arthur Cayley nurtured the theory of matrices and included the definition of matrix multiplication – he used the term *matrix composition*– in a paper of 1858, though that paper only considered 2×2 and 3×3 matrices.

The theory of matrices was developed in two quite different ways by its early pioneers. Sylvester and Cayley are examples of mathematicians who emphasized abstract algebraic structure. The Irish mathematician William Rowan Hamilton and German mathematician Herman Grassmann are examples of mathematicians who favoured a geometric view of matrices. The theory has benefited enormously from the complementarity of these two approaches.

12.1 Algebra versus Geometry

The preceding lectures have focused on algebraic aspects of matrices – their addition, multiplication, division and the relationships satisfied by these operations. Our next aim is to provide a geometric view of matrices. The following excerpt from a lecture delivered by Fields Medallist Michael Atiyah in 2000 explains the case for geometry.

> Let me try to explain my own view of the difference between geometry and algebra. Geometry is, of course, about space, of that there is no question. If I look out at the audience in this room I can see a lot; in one single second or microsecond I can take in a vast amount of information, and that is of course not an accident. Our brains have been constructed in such a way that they are extremely concerned with vision. Vision, I understand from friends who work in neurophysiology, uses up something like 80 or 90 percent of the cortex of the brain. There are about 17 different

centres in the brain, each of which is specialised in a different part of the process of vision: some parts are concerned with vertical, some parts with horizontal, some parts with colour, or perspective, and finally some parts are concerned with meaning and interpretation. Understanding, and making sense of, the world that we see is a very important part of our evolution. Therefore, spatial intuition or spatial perception is an enormously powerful tool, and that is why geometry is actually such a powerful part of mathematics—not only for things that are obviously geometrical, but even for things that are not. We try to put them into geometrical form because that enables us to use our intuition. Our intuition is our most powerful tool. That is quite clear if you try to explain a piece of mathematics to a student or a colleague. You have a long difficult argument, and finally the student understands. What does the student say? The student says, 'I see!' Seeing is synonymous with understanding, and we use the word 'perception' to mean both things as well. At least this is true of the English language. It would be interesting to compare this with other languages. I think it is very fundamental that the human mind has evolved with this enormous capacity to absorb a vast amount of information, by instantaneous visual action, and mathematics takes that and perfects it.

Algebra, on the other hand (and you may not have thought about it like this), is concerned essentially with time. Whatever kind of algebra you are doing, a sequence of operations is performed one after the other and 'one after the other' means you have got to have time. In a static universe you cannot imagine algebra, but geometry is essentially static. I can just sit here and see, and nothing may change, but I can still see. Algebra, however, is concerned with time, because you have operations which are performed sequentially and, when I say 'algebra', I do not just mean modern algebra. Any algorithm, any process for calculation, is a sequence of steps performed one after the other; the modern computer makes that quite clear. The modern computer takes its information in a stream of zeros and ones, and it gives the answer.

Algebra is concerned with manipulation in time and geometry is concerned with space. These are two orthogonal aspects of the world, and they represent two different points of view in mathematics. Thus the argument or dialogue between mathematicians in the past about the relative importance of geometry and algebra represents something very, very fundamental.

Of course it does not pay to think of this as an argument in which one side loses and the other side wins. I like to think of this in the form of an analogy: 'Should you just be an algebraist or a geometer?' is like saying 'Would you rather be deaf or blind?' If you are blind, you do not see space: if you are deaf, you do not hear, and hearing takes place in time. On the whole, we prefer to have both faculties.

In physics, there is an analogous, roughly parallel, division between the concepts of physics and the experiments. Physics has two parts to it: theory—concepts, ideas, words, laws—and experimental apparatus. I think that concepts are in some broad sense geometrical, since they are concerned with things taking place in the real world. An experiment, on the other hand, is more like an algebraic computation. You do something in time; you measure some numbers; you insert them into formulae, but the basic concepts behind the experiments are a part of the geometrical tradition.

One way to put the dichotomy in a more philosophical or literary framework is to say that algebra is to the geometer what you might call the 'Faustian offer'. As you know, Faust in Goethe's story was offered whatever he wanted (in his case the love of a beautiful woman), by the devil, in return for selling his soul. Algebra is the offer made by the devil to the mathematician. The devil says: ‘I will give you this powerful machine, it will answer any question you like. All you need to do is give me your soul: give up geometry and you will have this marvellous machine.’ (Nowadays you can think of it as a computer!) Of course we like to have things both ways; we would probably cheat on the devil, pretend we are selling our soul, and not give it away. Nevertheless, the danger to our soul is there, because when you pass over into algebraic calculation, essentially you stop thinking; you stop thinking geometrically, you stop thinking about the meaning.

I am a bit hard on the algebraists here, but fundamentally the purpose of algebra always was to produce a formula which one could put into a machine, turn a handle and get the answer. You took something that had a meaning; you converted it into a formula, and you got out the answer. In that process you do not need to think any more about what the different stages in the algebra correspond to in the geometry. You lose the insights, and this can be important at different stages. You must not give up the insight altogether! You might want to come back to it later on. That is what I mean by the Faustian offer. I am sure it is provocative.

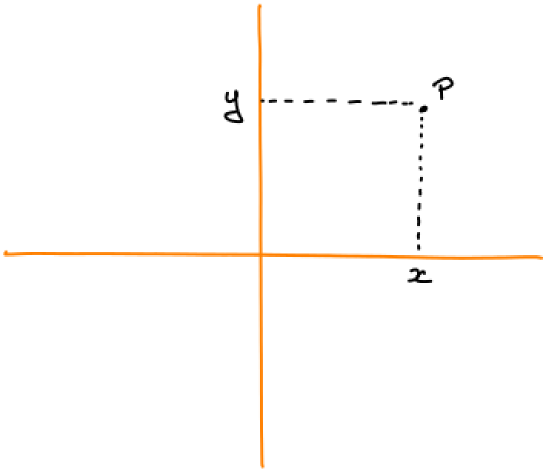
This choice between geometry and algebra has led to hybrids which confuse the two, and the division between algebra and geometry is not as straightforward and naïve as I just said. For example, algebraists frequently will use diagrams. What is a diagram except a concession to geometrical intuition?

12.2 Linear transformations

We refer to the set $\mathbb{R}$ of real numbers as the *real line* and picture it as an infinite line with no start or end points. We refer to the set

$$\mathbb{R}^2 = \{(x,y) : x,y \in \mathbb{R}\} \tag{12.1}$$

of pairs of real numbers as the *real plane*, or just *plane*, and picture it as an infinite plane with no boundary.



Any point P in the plane can be represented by a pair (x,y) of real numbers.

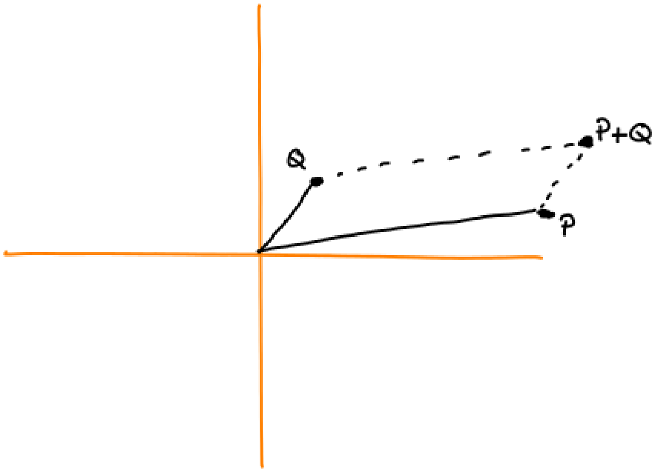
A *transformation* of the plane is just a function

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \tag{12.2}$$

which sends each point $P = (x,y)$ to some point $T(P)$ in the plane.

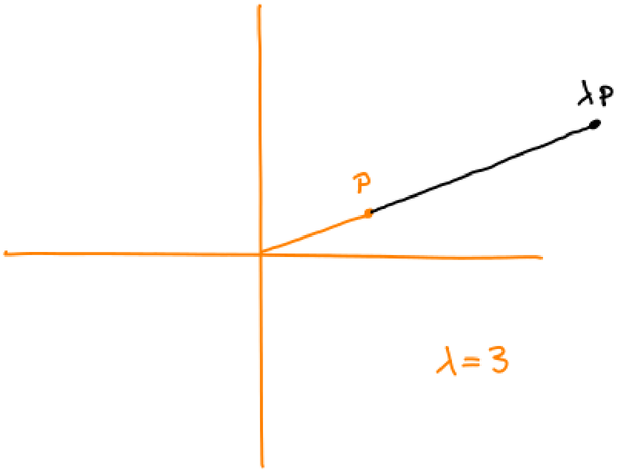
We can add two points $P = (x,y)$ and $Q = (x',y')$ using matrix addition of row vectors.

$P+Q=(x+x',y+y')$



We can multiply a point $P=(x,y)$ by a scalar $\lambda\in\mathbb{R}$ using scalar multiplication of matrices.

$\lambda P=(\lambda x,\lambda y)$



Definition 12.2.1 A transformation $T:\mathbb{R}^2\to\mathbb{R}^2$ is said to be *linear* if

1. $T(P+Q)=T(P)+T(Q)$
2. $T(\lambda P)=\lambda T(P)$

for all $P,Q\in\mathbb{R}^2$, $\lambda\in\mathbb{R}$.

■ **Example 12.1** Consider the transformation

$$T:\mathbb{R}^2\to\mathbb{R}^2,(x,y)\mapsto(3x+7y,2x+5y)\ . \tag{12.3}$$

So, for instance, $T(-3,1)=(-2,-1)$.

For arbitrary points $P = (x,y)$, $Q = x',y')$ we find:

$$\begin{aligned} T(P+Q) &= T(x+x',y+y') && (12.4)\\ &= (3(x+x')+7(y+y'),2(x+x')+5(y+y')) && (12.5)\\ &= (3x+7y+3x'+7y',2x+5y+2x'+5y') && (12.6)\\ &= (3x+7y,2x+5y)+(3x'+7y',2x'+5y') && (12.7)\\ &= T(P)+T(Q) && (12.8)\end{aligned}$$

Also, for any $\lambda \in \mathbb{R}$ we find:

$$\begin{aligned} T(\lambda P) &= T(\lambda x,\lambda y) && (12.9)\\ &= (3\lambda x+7\lambda y,2\lambda x+5\lambda y)) && (12.10)\\ &+\lambda(3x+7y,2x+5y) && (12.11)\\ &= \lambda T(P) && (12.12)\end{aligned}$$

Thus T is a linear transformation. ■

■ **Example 12.2** Consider the transformation

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2, (x,y) \mapsto (x^2,y^2)\ . \qquad (12.13)$$

So, for instance, $T(-3,1) = (9,1)$.

For $P = (1,2)$ and $\lambda = 3$ we find:

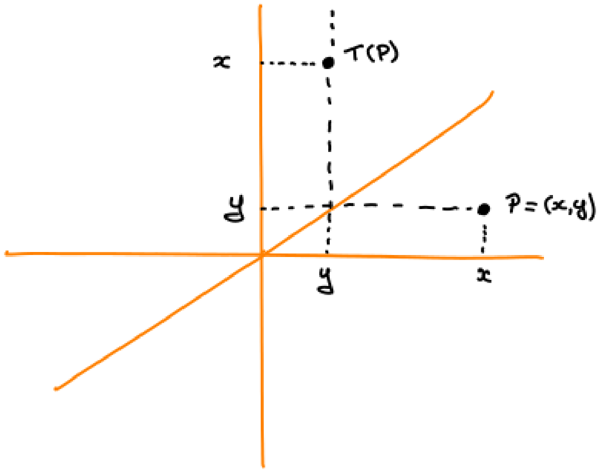
$$\begin{aligned} T(\lambda P) &= T(3,6) && (12.14)\\ &= (9,36) && (12.15)\end{aligned}$$

$$\begin{aligned} \lambda T(P) &= 3T(1,2) && (12.16)\\ &= 3(1,4) && (12.17)\\ &= (3,12) && (12.18)\end{aligned}$$

Since $T(\lambda P) \neq \lambda T(P)$ in this particular case, we conclude that T is not a linear transformation. ■

■ **Example 12.3** Let : $\mathbb{R}^2 \to \mathbb{R}^2$ be the transformation of the plane obtained by reflecting in the line $y = x$. To decide whether T is linear we need to find an algebraic formula for the transformation.

From the diagram



we find:

$$T(x,y) = (y,x) \tag{12.19}$$

For arbitrary points $P = (x,y), Q = (x',y')$ and scalar $\lambda \in \mathbb{R}$ we find:

$$\begin{aligned} T(P+Q) &= T(x+x',y+y') && (12.20)\\ &= (y+y',x+x') && (12.21)\\ &= (y,x)+(y',x') && (12.22)\\ &T(P)+T(Q) && (12.23) \end{aligned}$$

$$\begin{aligned} T(\lambda P) &= T(\lambda x, \lambda y) && (12.24)\\ &= (\lambda y, \lambda x) && (12.25)\\ &= \lambda(y,x) && (12.26)\\ &\lambda T(P) && (12.27) \end{aligned}$$

Hence reflection in the line $y = x$ is a linear transformation. ■

12.3 Transformations that preserve lines

Let P and V be row vectors in $\mathbb{R}^2$. Any set of the form

$$L = \{P + \lambda V \ : \ \lambda \in \mathbb{R}\} \tag{12.28}$$

is called a *line*. We think of L as the collection of points in the plane that 'can be reached by starting at P and travelling some distance in the direction of V'. Note that we allow the case $V = 0$ in which the line L consists of just a single point.

A transformation of the plane $T: \mathbb{R}^2 \to \mathbb{R}^2$ is said to be *line preserving* if the set

$$T(L) = \{T(v): \ v \in L\} \tag{12.29}$$

is a line whenever L is a line. Since a linear transformation T satisfies $T(P+\lambda V) = T(P) + \lambda T(V)$ we immediately have the following result.

Proposition 12.3.1 Any linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ is line preserving.

So linear transformations do indeed have something to do with lines. However, not every line preserving transformation is linear!

13. Geometry of matrices

The linear transformation

$$T : \mathbb{R}^2 \to \mathbb{R}^2, (x,y) \mapsto (3x+7y, 2x+5y) \tag{13.1}$$

can be represented using matrix notation and the rules for matrix multiplication as follows.

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 3 & 7 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3x+7y \\ 2x+3y \end{pmatrix} \tag{13.2}$$

We say that the matrix

$$A = \begin{pmatrix} 3 & 7 \\ 2 & 5 \end{pmatrix} \tag{13.3}$$

represents the linear transformation T.

In fact, the distributivity rule (9.5) for matrix multiplication

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \end{pmatrix} \tag{13.4}$$

is a linear transformation. Thus *any* 2×2 matrix A represents a linear transformation of the plane in this way. The following result shows that the converse also holds.

Theorem 13.0.1 Any linear transformation of the plane

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$

can be represented by a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

of real numbers.

Proof. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be an arbotrary linear transformation.

$$v = \begin{pmatrix} x \\ y \end{pmatrix} \tag{13.5}$$

denote an arbitrary (column) vector in $\mathbb{R}^2$. Consider the two articular vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{13.6}$$

Now $T(e_1)$ and $T(e_2)$ have some values, say

$$T(e_1) = \begin{pmatrix} a \\ c \end{pmatrix}, \quad T(e_2) = \begin{pmatrix} b \\ d \end{pmatrix}. \tag{13.7}$$

Using (13.7) and the linearity of T we find:

$$\begin{aligned}
T(v) &= T\begin{pmatrix} x \\ y \end{pmatrix} && (13.8)\\
&= T(x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}) && (13.9)\\
&= T(xe_1 + ye_2) && (13.10)\\
&= xT(e_1) + yT(e_1) \qquad \text{(by the linearity of } T\text{)} && (13.11)\\
&= x\begin{pmatrix} a \\ c \end{pmatrix} + y\begin{pmatrix} b \\ d \end{pmatrix} && (13.12)\\
&= \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix} && (13.13)\\
&= \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. && (13.14)
\end{aligned}$$

Hence T is represented by a 2×2 matrix of real numbers. ■

■ **Example 13.1** A linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ satisfies $T(1,2) = (1,3)$ and $T(3,4) = (4,5)$. Find a formula for $T(x,y)$.

We have:

$$\begin{aligned}
T(1,0) &= T(2(1,2) - (3,4)) && (13.15)\\
&= 2T(1,2) - T(3,4) && (13.16)\\
&= (2,6) - (4,5) && (13.17)\\
&= (-2,1) && (13.18)
\end{aligned}$$

$$\begin{aligned}
T(0,1) &= T(-\frac{3}{2}(1,2) + \frac{1}{2}(3,4)) && (13.19)\\
&= -\frac{3}{2}T(1,2) + \frac{1}{2}T(3,4) && (13.20)\\
&= -\frac{3}{2}(1,3) + \frac{1}{2}(4,5) && (13.21)\\
&= (\frac{1}{2}, -2) && (13.22)
\end{aligned}$$

Hence

$$T(x,y) = x(-2,1) + y(\frac{1}{2},-2) = (-2x + \frac{1}{2}y, x - 2y) \tag{13.23}$$

is the required formula. ■

13.1 Why did that proof work?

Having read the proof of a theorem line by line, making sure that each line follows from previous lines, we are in a position to say "Ah, now I know for sure that the theorem is true". But a line by line reading may not be enough for us to say "Ah, now I really see why the result is true". To "really see" why a result is true, we might need to develop an informative overview of the proof.

The proof of Theorem 13.0.1 has two main ingredients: i) linearity of T is fundamental to the argument; ii) the choice of the two vectors e_1 and e_2 is the other key ingredient. The two vectors e_1 and e_2 have the property that for any vector $v \in \mathbb{R}^2$ there is a unique pair of real numbers x, y such that $v = xe_1 + ye_2$. In fact, this is the only property of e_1 and e_2 that was used. So in the proof we could have taken *any* two vectors with this property. A pair of vectors with this property is called a *basis* of $\mathbb{R}^2$. From the geometric picture for addition of vectors presented in the previous lecture, it is clear that any pair of vectors will do providing that the rays from the origin to each of them them are not colinear. The particular pair e_1, e_2 used in the proof is said to be the *standard basis* of $\mathbb{R}^2$.

Now that we can see why Theorem 13.0.1 is true, it is routine to generalize its ingredients to the following definition and result about $n \times n$ matrices.

Definition 13.1.1 A transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *linear* if

1. $T(P+Q) = T(P) + T(Q)$
2. $T(\lambda P) = \lambda T(P)$

for all $P, Q \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$.

Theorem 13.1.1 Any linear transformation

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

can be represented as

$$T : \mathbb{R}^n \to \mathbb{R}^n, v \mapsto Av$$

where A is an $n \times n$ matrix.

The proof of Theorem 13.1.1 uses the following definition.

Definition 13.1.2 A list of n vectors $e_1, e_2, \ldots, e_n \in \mathbb{R}^n$ is said to be a basis for $\mathbb{R}^n$ if, for any vector $v \in \mathbb{R}^n$, there is a unique list of real numbers $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that

$$v = x_1e_1 + x_2e_2 + \cdots + x_ne_n \ .$$

■ **Example 13.2** The vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{13.24}$$

form a basis for $\mathbb{R}^3$. This particular basis is called the *standard basis* for $\mathbb{R}^3$. ■

13.2 Matrix multiplication explained

On first encounter the formula (9.11) for matrix multiplication may seem puzzling. It may seem like pure chance, or perhaps clever ingenuity on the part of its inventor, that this multiplication satisfies familiar properties such as associativity (9.5) and distibutivity over addition (9.5). However, the following theorem shows that formula (9.11) is immediately stumbled upon by anyone wishing to calculate with composites of linear transformations.

Theorem 13.2.1 Let $S\colon \mathbb{R}^n \to R^n, v \mapsto Av$ and $T\colon \mathbb{R}^n \to R^n, v \mapsto Bv$ be two linear transformations represented by $n \times n$ matrices A and B respectively. Then the composite function

$$S \circ T\colon \mathbb{R}^n \to R^n, v \mapsto S(T(v)) \tag{13.25}$$

satisfies

$$S \circ T(v) = (AB)v\ . \tag{13.26}$$

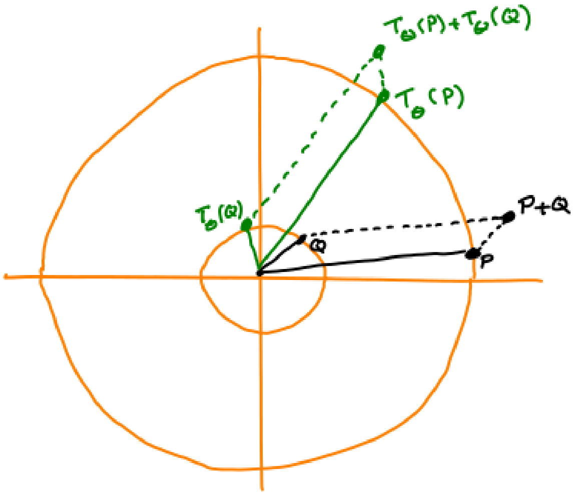
To prove this theorem one just needs to verify the equality (13.26). It is a straightforward and worthwile excercise to verify (13.26) for $n = 2, 3$, after which it will be clear that the equality holds for all $n \geq 1$.

Since the product AB of two $n \times n$ matrices is itself an $n \times n$ matrix, and since multiplication by a matrix represents a linear transformation, we arrive at the following important consequence of Theorem 13.2.1.

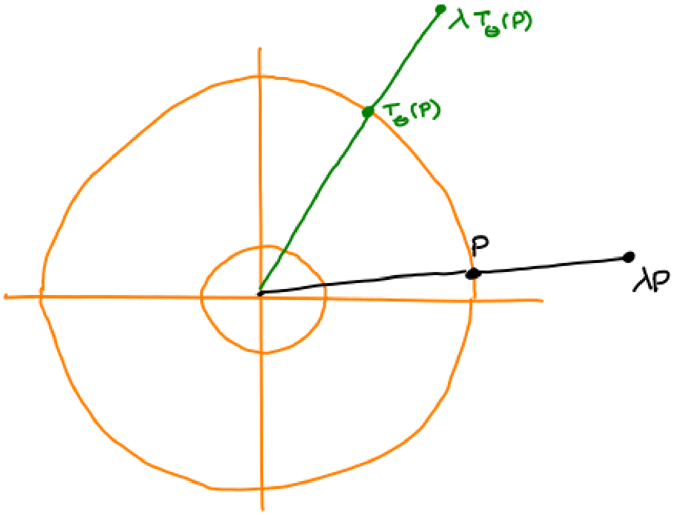
Corollary 13.2.2 The composite $S \circ T$ of two linear transformations of $\mathbb{R}^n$ is itself a linear transformation.

13.3 Rotations and reflections

Let $T_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote the transformation that rotates the real plane anticlockwise through an angle θ about the origin $(0,0)$. The following diagram pictures two parallelograms, one rotated clockwise through an angle θ. One parallelogram represents the addition of two points P, Q. The other parallelogram represents the addition of $T_\theta(P)$, $T_\theta(Q)$.



From this diagram we see that $T_\theta(P+Q) = T_\theta(P) + T_\theta(Q)$. The next diagram pictures a point P, a scalar multiple of it λP, the point $T_\theta(P)$ and the scalar multiple $\lambda T_\theta(P)$.



We see that $T_\theta(\lambda P) = \lambda T_\theta(P)$. So we have proved the following.

Theorem 13.3.1 A rotation $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ of the plane about the origin is a linear transformation.

In Example 12.3 we proved that reflection in the line $y = x$ is a linear transformation of the plane. If we take an arbitrary line through the origin then reflection in this arbitrary line is the same as an anticlockwise rotation about the origin through an angle θ so that the arbitrary line coincides with the line $y = x$, followed by reflection in the line $y = x$, followed by a clockwise rotation through θ. The following theorem follows from Corollary 13.2.2.

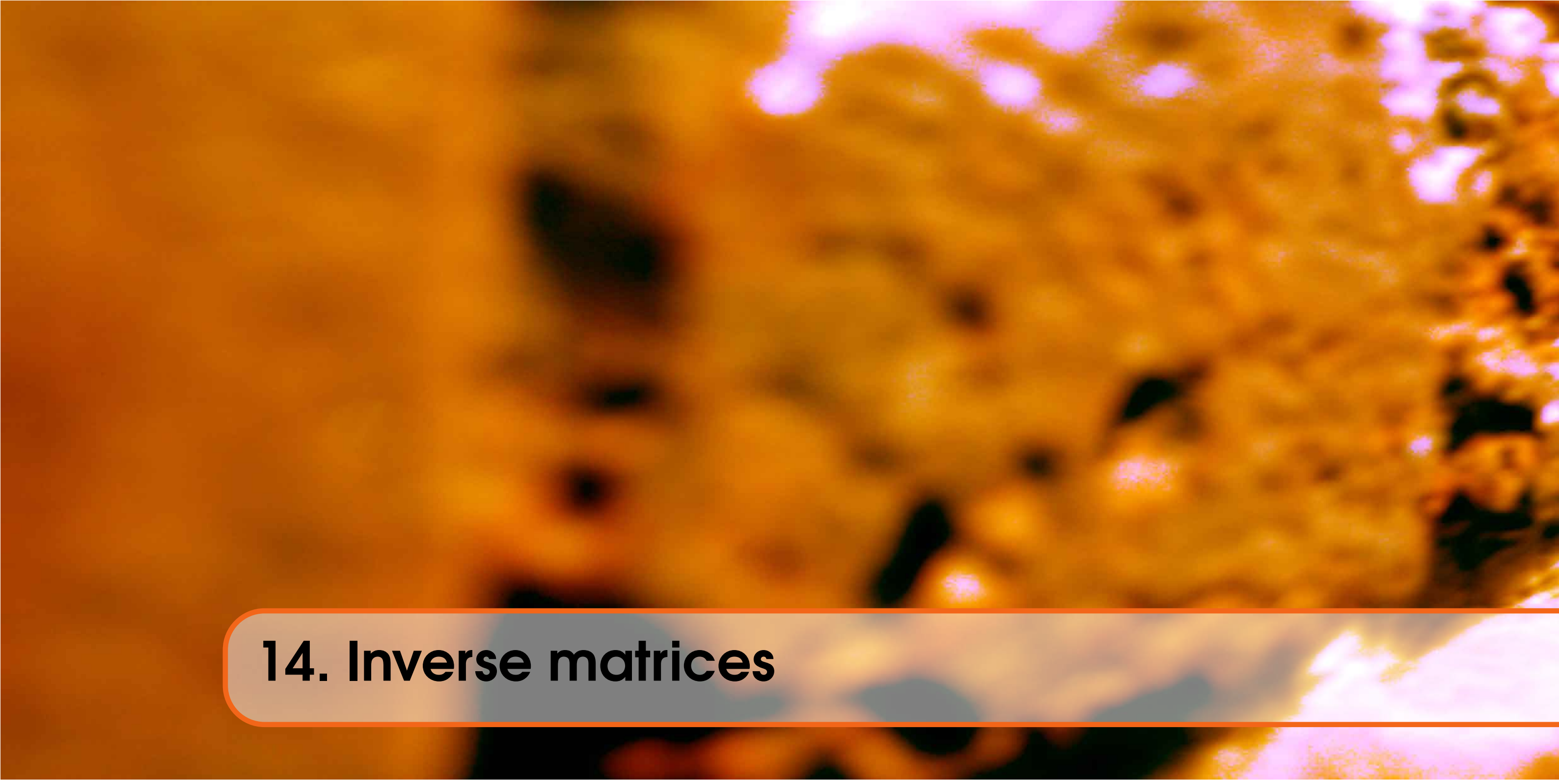
Theorem 13.3.2 A reflection $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ of the plane in a line through the origin is a linear transformation.

■ **Example 13.3** Find a formula for the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ consisting of reflection in the y-axis followed by clockwise rotation about the origin through an angle of $5\pi/2$ radians.

To find a formula we first note that the reflection and rotation are both linear transformations. Hence their composite T is a linear transformation. A formula can thus be derived from the values of $T(1,0)$ and $T(0,1)$. By inspection, $T(1,0) = (0,1)$ and $T(0,1) = (1,0)$. The required formula is thus:

$$T(x,y) = (y,x) \tag{13.27}$$

Note that this formula tells us that the composite transformation T is simply the transformation that reflects in the line $y = x$. ■



14. Inverse matrices

Consider the two matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 5 \\ 3 & 8 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 10 & -12 & 5 \\ -3 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix}. \tag{14.1}$$

We can multiply them to find:

$$AB = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I. \tag{14.2}$$

From equation 14.2 we deduce that the inverse of matrix A is

$$A^{-1} = B. \tag{14.3}$$

Two questions may come to mind: (i) how would we have found B had it not been given; (ii) does anybody really care about finding the inverse of A?

As a partial answer to Question (ii), consider the following system of *linear* equations.

$$\begin{array}{rcrcrcl} x & + & 2y & + & 3z & = & 1 \\ 2x & + & 5y & + & 5z & = & 2 \\ 3x & + & 8y & + & 6z & = & 3 \quad . \end{array} \tag{14.4}$$

These equations are said to be *linear* because the unkown quantities x, y, z don't appear as powers such as x^2, y^3, z^{-4} nor as products such as xy or xyz. This system of equations can be re-expressed as follows, using matrix multiplication.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 5 \\ 3 & 8 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{14.5}$$

To determine the values of x, y, z satisfying (14.4) we can multiply both sides of (14.5) by A^{-1} to obtain

$$AA^{-1}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^{-1}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{14.6}$$

and thus, from the above value of A^{-1} ,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 10 & -12 & 5 \\ -3 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{14.7}$$

We have illustrated how an inverse matrix can be used to find the solution $x = 1, y = 0, z = 0$ to a system of linear equations. Systems of linear equations abound in the sciences, social sciences, and engineering; opportunities for using inverse matrices also abound.

We now turn to Question (i): given a matrix A, how should we go about finding an inverse A^{-1}. We provide one of several answers to this.

14.1 Gauss-Jordan method for finding the inverse of an invertible matrix

Suppose that we wish to find the inverse of some $n \times n$ matrix A. The entries of A could be real numbers or they could be numbers in clock arithmetic. To cover both possibilities we let $\mathbb{K}$ denote the real numbers $\mathbb{R}$ or the integers $\mathbb{Z}_N$ modulo some positive integer N, and say that A is *over* $\mathbb{K}$. To find the inverse of A we can first use the $n \times n$ identity matrix I to form an $n \times 2n$ matrix $(A \mid I)$. We can then try to apply a sequence of suitably defined *elementary row operations*

$$(A \mid I) \quad \xrightarrow{row-ops} \quad (I \mid B) \tag{14.8}$$

to transform the $n \times 2n$-matrix to one of the form $(I \mid B)$ where B is some $n \times n$-matrix over $\mathbb{K}$. The definition of the elementary operations will ensure that $B = A^{-1}$. There are three allowable row operations:

(I) $R_i \leftarrow R_i + \lambda R_j \quad j \neq i, \lambda \in \mathbb{K}$.
Add a multiple of the jth row to the ith row.

(II) $R_i \leftarrow \lambda R_i \quad \lambda \in \mathbb{R}$, λ invrtible.
Multiply the ith row by an invertible number.

(III) $R_i \leftrightarrow R_j$.
Interchange the ith and jth rows.

To illustrate the method we use it to find the inverse of the above matrix A.

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 5 & 0 & 1 & 0 \\ 3 & 8 & 6 & 0 & 0 & 1 \end{array}\right) \tag{14.9}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -3 & 0 & 1 \end{array}\right) \begin{array}{l} R_2 \leftarrow R_2 - 2R_1, \\ R_3 \leftarrow R_3 - 3R_1 \end{array} \tag{14.10}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right) R_3 \leftarrow R_3 - 2R_2 \tag{14.11}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right) R_3 \leftarrow -R_3 \tag{14.12}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 4 & -6 & 3 \\ 0 & 1 & 0 & -3 & 3 & -1 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right) \begin{array}{l} R_1 \leftarrow R_1 - 3R_3 \\ R_2 \leftarrow R_2 + R_3 \end{array} \tag{14.13}$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -12 & 5 \\ 0 & 1 & 0 & -3 & 3 & -1 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right) R_1 \leftarrow R_1 - 2R_2 \tag{14.14}$$

We conclude that

$$A^{-1} = \begin{pmatrix} 4 & -12 & 5 \\ -3 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix}. \tag{14.15}$$

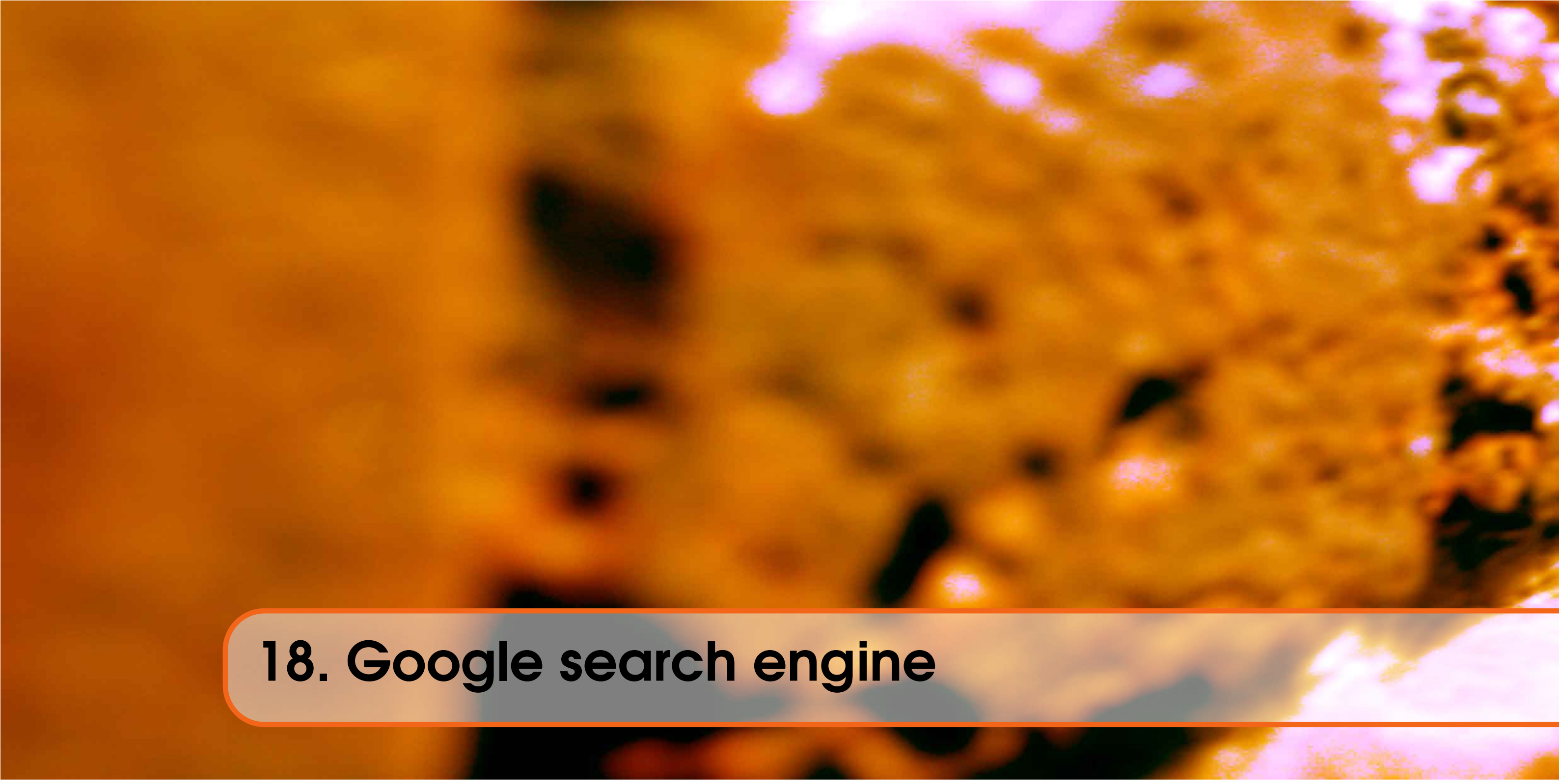
In the next lecture we'll explain why this process always furnishes the inverse of an invertible matrix.

15. Row operations

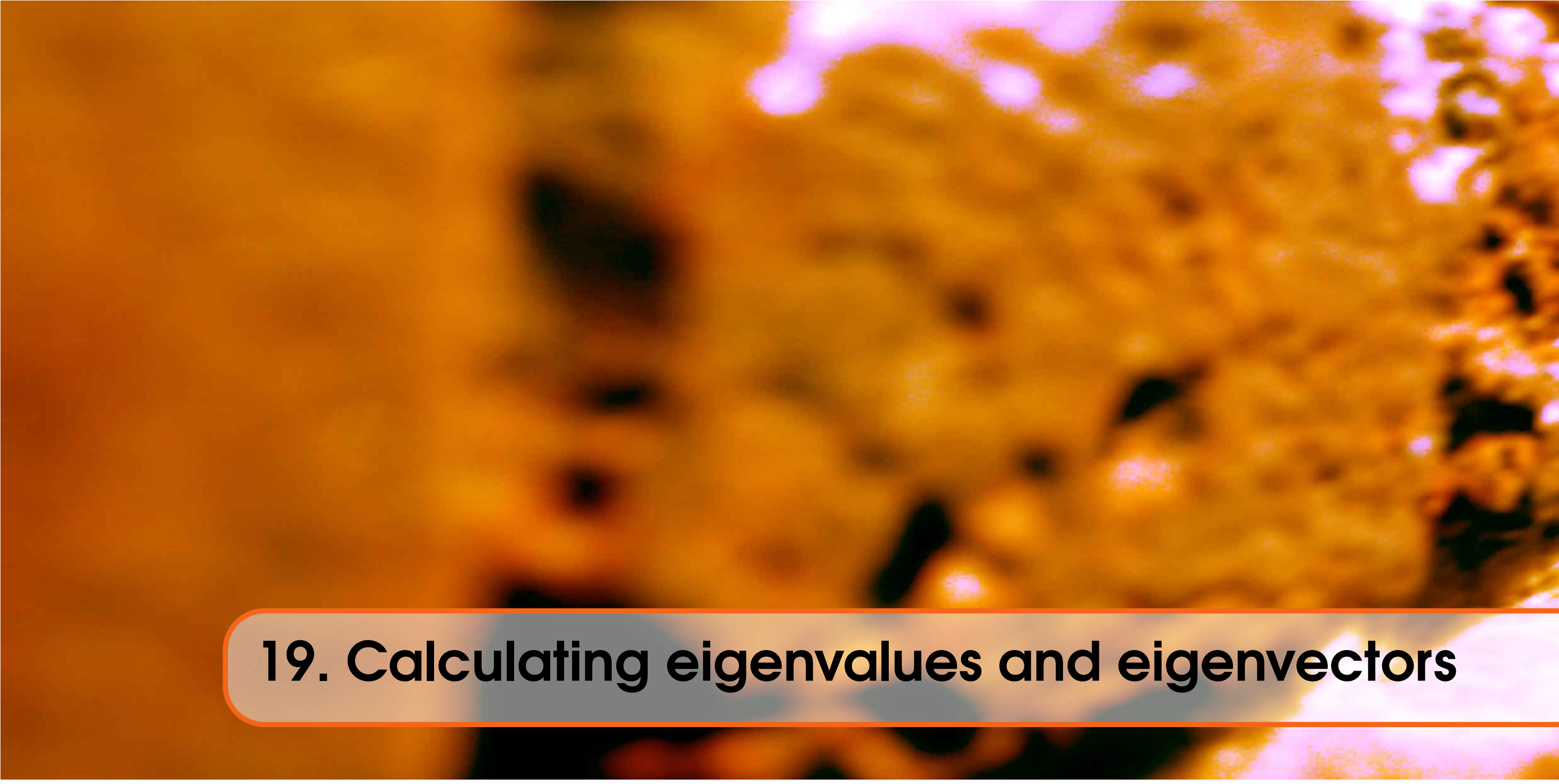
16. Determinants

III Eigenvalues

17. Eigenvalues and eigenvectors



18. Google search engine

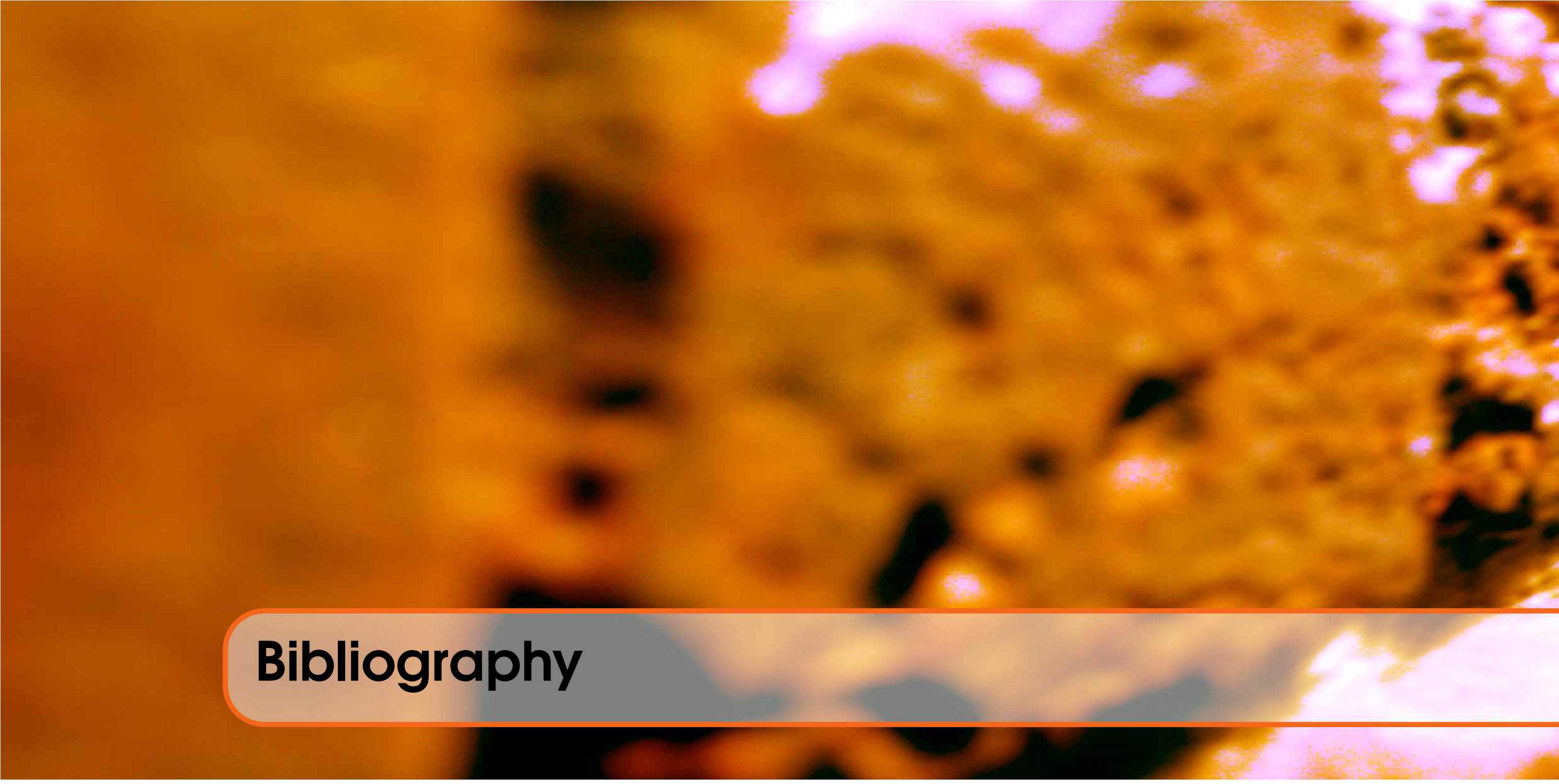


19. Calculating eigenvalues and eigenvectors

20. The Golden Ratio

21. More on the Golden Ratio

22. A Model of infectious diseases



Bibliography