Advanced Mathematics
Theory, notes and examples

Engineering Informatics

Gerardo Oleaga

January 25, 2016

Contents

I. Calculus **4**

1. Sequences and series of functions **5**
1.1. Review of sequences and limits of numbers 5
1.2. Sequences of functions 7
1.3. Series of numbers 9
1.4. Series of functions 14

2. Fourier Series and Transforms **16**
2.1. Fourier series 16
2.2. Computing Fourier coefficients: examples 19
2.3. Quadratic approximation 24
2.4. Fourier transform 28
2.5. Application to signal filtering 30

3. First Order Differential Equations **33**
3.1. Introduction 33
3.2. Variable separation 36
3.3. The first order linear equation 38

4. Second order linear equations **43**
4.1. Introduction 43
4.2. The homogeneous equation with constant coefficients 45
4.3. The non-homogeneous equation with constant coefficients 47
4.4. Laplace transform 50

Bibliography **54**

II. Algebra **55**

5. Integers **56**
5.1. The induction principle 56
5.2. Primes and factorization 60
5.3. Linear diophantine equations 62
5.4. Congruences 64
5.5. Congruences and powers 71

6. Groups **77**
6.1. Definitions and general properties 77
6.2. Permutation Groups 79
6.3. Group morphisms and quotient groups 79
6.4. Finite Groups 81

7. Rings and Fields **87**
7.1. Definitions and general properties 87
7.2. Ring classification 87
7.3. Integral Domains and Fields 89
7.4. Unique Factorization Domains 91

8. Polynomials **92**
8.1. Polynomials 93
8.2. The division algorithm 94
8.3. Greatest common divisor, Bézout identity and Euclid's algorithm . . . 96
8.4. Irreducible polynomials and factorization 98
8.5. Ideals and Congruences in $K[x]$ 101
8.6. If K is a finite field 105

Bibliography **106**

Part I.

Calculus

1. Sequences and series of functions

Bibliography: *Applied Mathematics: Body & Soul.* Kenneth Eriksson, Don Estep and Claes Johnson, Vol 1. Available on the web: http://www.csc.kth.se/~cgjoh/eriksson-vol1.pdf.

1.1. Review of sequences and limits of numbers

Rational numbers in decimal representation are a good example of the concepts of sequences and limits.

$$\begin{aligned} 10/9 &= 1 + 1/9 \\ &= 1 + (10/9) \times 10^{-1} \,. \end{aligned}$$

Now we may continue recursively:

$$\begin{aligned} 10/9 &= 1 + (1 + 1/9) \times 10^{-1} \\ &= 1 + \left(1 + (10/9) \times 10^{-1}\right) \times 10^{-1} \\ &= 1.1 + (10/9) \times 10^{-2} \,. \end{aligned}$$

Following the same procedure n times we find:

$$10/9 = 1.\underbrace{1111\ldots1}_{n} + (10/9) \times 10^{-(n+1)} \,.$$

An alternative way to show this is that we may approximate the fraction $10/9$ with a finite quantity n of decimal digits:

$$\left| 10/9 - 1, \underbrace{1111\ldots1}_{n} \right| = \frac{1}{9 \times 10^n} \,.$$

Increasing the number of digits, we see that the decimal representation *converges* to $10/9$, because the *distance* between the representation and the fraction goes to zero like $1/(9 \times 10^n)$ in terms of n, that is, the number of digits included. The quantities:

$$\begin{aligned} a_0 &= 1 \\ a_1 &= 1.1 \\ a_2 &= 1.11 \\ &\vdots \\ a_n &= 1.1111\ldots \end{aligned}$$

are a *sequence* of real numbers (in this case, *rational numbers*) that approximate the number 10/9. To indicate that a_n is *arbitrarily close* to 10/9, we write:

$$\lim_{n\to\infty} a_n = 10/9\,.$$

That the sequence of numbers *has a limit* L, means that we can approach arbitrarily the number L if we choose n big enough. The mathematical expression for such a sequence is:

$$\{a_n\}_{n=1}^{\infty}$$

and when it is convergent to a number L we write:

$$\lim_{n\to\infty} a_n = L\,.$$

Remark: The notion of convergence must be more precise to avoid unwanted cases, such as a sequence that is arbitrarily close to *two* different numbers.

Example: consider the sequence

$$a_n = (-1)^n + \frac{1}{n}\,.$$

These numbers are (alternatively) very close to number 1, but also close to -1. When n is even, that is, if $n = 2k$, for $k = 0, \ldots \infty$, we have that:

$$|1 - a_{2k}| = \frac{1}{n}\,.$$

Moreover, when n is odd, that is $n = 2k + 1$, we have that:

$$|-1 - a_{2k+1}| = \frac{1}{n}\,.$$

As we showed in the example, a sequence may attain values arbitrarily close to several numbers. When a number is given in decimal form we want to give a reference to a *unique value.* We must precise a bit more the definition of the limit of a sequence. We will ask that *every value from one index N (to be determined) must be close to the limit.* The number N will depend on the distance that we require to the limit value.

Definition *Limit of a sequence of numbers.* The concept of limit was made precise by Karl Wierstrass (1815–97). The sequence $\{a_n\}_{n=1}^{\infty}$ has a limit L if and only if for every number $\varepsilon > 0$ (arbitrarily small), there exists a natural number N_ε (dependent on epsilon) such that

$$n > N_\varepsilon \Rightarrow |a_n - L| < \varepsilon\,.$$

The definition gives no clue about the quantitative relationship between ε and N_ε, it only guarantees the existence of such a relationship.

Some useful properties to compute limits

We assume that the sequences $\{a_n\}_{n=1}^{\infty}$ y $\{b_n\}_{n=1}^{\infty}$ have limits A and B respectively.

1. $\lim_{n\to\infty}(a_n + b_n) = A + B$.
2. $\lim_{n\to\infty}(a_n \times b_n) = A \times B$.
3. If f is a *continuous* function on $x = A$, then $\lim_{n\to\infty} f(a_n) = f(\lim_{n\to\infty} a_n) = f(A)$.

1.2. Sequences of functions

Bibliography: *Introduction to Real Analysis*, Robert G. Bartle y Donald Sherbert. John Wiley & Sons, Inc. 2011 (4th edition).

Given a set $D \subset \mathbb{R}$, a sequence of functions is a family of functions defined on D indexed by natural numbers: For each fixed x, $\{f_n(x)\}_{n=0}^{\infty}$ is a sequence of real or complex numbers.

Pointwise convergence

Definition We say that the function sequence $\{f_n\}_{n=0}^{\infty}$ is *pointwise convergent* on D to a function f if and only if for each $x \in D$, the sequence $\{f_n(x)\}_{n=0}^{\infty}$ converges to a (real or complex) number, denoted by $f(x)$, that is:

$$f(x) := \lim_{n\to\infty} f_n(x) \ .$$

Using this definition we can say that the sequence of functions f_n is pointwise convergent to the function f if and only if for each $x \in D$ y and for each $\varepsilon > 0$, there exists a natural number $N_{x,\varepsilon}$ such that, for $n > N_{x,\varepsilon}$, then $|f_n(x) - f(x)| < \varepsilon$.

Uniform convergence

Uniform convergence differs in a subtle way from pointwise convergence. The formal definition is given by:

Definition The sequence f_n is uniformly convergent to a function f on the domain D if for each $\varepsilon > 0$ there exists a natural number N_ε (depending only on ε) such that, for every $x \in D$:

$$|f_n(x) - f(x)| < \varepsilon \, .$$

Cauchy criterion for uniform convergence of functions: f_n converges uniformly to f if, given $\varepsilon > 0$, there exists N_ε such that *for all* $x \in D$

$$m, n > N_\varepsilon \Rightarrow |f_m(x) - f_n(x)| < \varepsilon \, .$$

We consider now a useful result to check that a sequence is *not uniformly convergent.*

Lemma A sequence of functions f_n does not converge uniformly to a function f on the domain D if and only if there exists some $\varepsilon > 0$ such that for all natural N we can find an $n > N$ and an $x \in D$ such that:

$$|f_n(x) - f(x)| > \varepsilon .$$

Theorem Let f_n be a sequence of continuous functions on a domain D that is uniformly convergent to a limit function denoted by f. Then f is *continuous.*

PROOF To show that f is continuous, we check the definition. Let us fix a point $x_0 \in D$. Given $\varepsilon > 0$ we must show that there exists a number $\delta > 0$, small enough, such that, if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \varepsilon$. Using uniform convergence, we choose then $\varepsilon > 0$ and n in such a way that

$$|f_n(x) - f(x)| < \varepsilon/3 \quad \forall x \in D .$$

By the continuity of f_n, given $\varepsilon/3$, there exists $\delta > 0$ such that, if $|x - x_0| < \delta$ then $|f_n(x) - f_n(x_0)| < \varepsilon/3$. We have then that, given $\varepsilon > 0$, there exists $\delta > 0$ such that, if $|x - x_0| < \delta$:

$$|f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \varepsilon .$$

In other words, the first term is small because of pointwise convergence in x, the second by the continuity of f_n and the third again because of the pointwise convergence in x_0. Due to the uniformity there is no need to choose different n's in x and x_0. That is the main part of the argument ■

Remark In other words: a continuous function cannot be arbitrarily and uniformly close to a discontinuous function.

Convergence of derivatives and integrals

Theorem If the sequence $\{f_n\}$ is uniformly convergent to the function f on the interval $[a, b] \subset \mathbb{R}$, and assuming that the integrals $I_n = \int_a^b f_n(x)\,\mathrm{d}x$ and $I = \int_a^b f(x)\,\mathrm{d}x$ exist, then $\lim_{n\to\infty} I_n = I$.

PROOF Que have to show that I is the limit of the sequence $\{I_n\}$. To this end, we consider the difference:

$$\begin{aligned} I_n - I &= \int_a^b f_n(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x \\ &= \int_a^b (f_n(x) - f(x))\,\mathrm{d}x \end{aligned}$$

so that

$$|I_n - I| \leq \int_a^b |f_n(x) - f(x)|\,\mathrm{d}x .$$

Uniform convergence allows to select N_ε such that, if $n > N_\varepsilon$, it holds that $|f_n(x) - f(x)| < \varepsilon$ for all x in $[a,b]$. In this case:

$$|I_n - I| \leq \int_a^b |f_n(x) - f(x)| \, \mathrm{d}x < \int_a^b \varepsilon \mathrm{d}x = \varepsilon (b-a) \; .$$

The number on the right hand side can be arbitrarily small by selecting ε small enough. Moreover, we can choose $\bar{\varepsilon} = \varepsilon (b-a)$ (arbitrarily small) and taking $N = N_\varepsilon$ we conclude that $|I_n - I| < \bar{\varepsilon}$ for all $n > N_\varepsilon$. ■

Theorem Let us assume that the sequence of functions $\{f_n\}$ converge pointwise to f in some interval $[a,b] \subset \mathbb{R}$. If the derivatives of f_n, denoted by f_n', exist, are continuous on the interval, and converge uniformly to a function g, then $f' = g$.

Proof Given $x \in [a,b]$, due to the continuity of the derivatives of f_n, we have that

$$f_n(x) - f_n(a) = \int_a^x f_n'(s) \, \mathrm{d}s \, .$$

If we take limit on both sides, due to the uniform convergence of f_n' to g (that is also continuous, because it is the uniform limit of continuous functions), and using the previous theorem on the interval $[a,x]$, we can show that:

$$\lim_{n\to\infty} \int_a^x f_n'(s) \, \mathrm{d}s = \int_a^x g(s) \, \mathrm{d}s \, .$$

On the other hand, we have that the f_n are pointwise convergent to f, then we conclude that:

$$f(x) - f(a) = \int_a^x g(s) \, \mathrm{d}s \, .$$

Applying the Fundamental Theorem of Calculus we conclude that:

$$f'(x) = g(x) = \lim_{n\to\infty} f_n'(x) \; .$$

■

1.3. Series of numbers

The concept of an infinite series has a central role in Calculus. A basic idea is the representation of arbitrary functions in terms of a sum of simple expressions. The representation of a real number is a basic example of the sum of a series:

$$\begin{aligned} s_0 &= 1 \\ s_1 &= 1 + \frac{1}{10} \\ s_3 &= 1 + \frac{1}{10} + \frac{1}{10^2} \end{aligned}$$

In this case we are adding smaller terms in each step. We have that:

$$s_n = 1,\ \underbrace{1\ldots1}_{n \text{ times}}$$

and we can show that:

$$\lim_{n\to\infty} s_n = \frac{10}{9}\,.$$

In a more general context, the same idea was applied by Fourier to represent functions by means of *trigonometric series* and by Wierstrass using *power series* or polynomials. In this section we make a short review of the most important properties of numerical series, leaving for the next chapter the trigonometric or Fourier series.

One of the simpler examples of a series is the *geometric series:*

$$s_n = 1 + a + a^2 + \cdots + a^n$$

where a is a real or complex number. This *partial sum* has a simple algebraic expression taking into account that

$$as_n = a + a^2 + \cdots + a^{n+1} = s_n - 1 + a^{n+1}$$

and solving for s_n we have that

$$s_n = \begin{cases} \frac{a^{n+1}-1}{a-1} & a \neq 1 \\ n+1 & a = 1 \end{cases}\,.$$

If $|a| \geq 1$, s_n diverges because a^{n+1} as well as $n+1$ have no limit. If $|a| < 1$ we have that $a^{n+1} \to 0$ and we find the value:

$$\lim_{n\to\infty} \sum_{k=0}^{n} a^k = \frac{1}{1-a} \quad |a| < 1\,.$$

A general numerical series is defined by the partial sums:

$$s_n = \sum_{k=0}^{n} a_k\,,$$

where $\{a_k\}_{k=0}^{\infty}$ is a sequence of real or complex numbers. We say that the series is *convergent* if and only if the limit of the partial sums $\{s_n\}_{n=0}^{\infty}$ exists. Therefore, every question about convergence of a numerical series can be reduced to the convergence of a sequence.

Necessary condition for convergence Consider the difference between two consecutive terms

$$s_n - s_{n-1} = a_n\,.$$

If $\lim_{n\to\infty} s_n = S$, then

$$\lim_{n\to\infty} (s_n - s_{n-1}) = \lim_{n\to\infty} a_n = 0\,.$$

In other words, *the general term of a convergent series must go to zero.* In practice, if we observe that the general term does not converge to zero, we conclude that the series cannot converge. In the geometric series, for example, when $|a| \geq 1$ the general term a^n does not converge to zero. However, if the general term goes to zero, *that does not mean that the series is convergent.* One typical example of this case is the series whose general term is $a_n = \frac{1}{n}$, that is divergent. The only thing we can say is that, when the general term does not converge to zero, then the series cannot be convergent. In other words, the condition $a_n \to 0$ *is not enough* for convergence. However, when the general term in the series *alternate signs* (and satisfies other conditions), we do have convergence.

We consider in the first place series with positive terms. In fact, given an arbitrary series there is another series of positive terms connected to it, defined by its *absolute values.* If the general term of a series is a_n, we have that, using the *triangular inequality*:

$$\left| \sum_{k=0}^{n} a_k \right| \leq \sum_{k=0}^{n} |a_k| \ ,$$

where the series of the right has positive terms.

Series with non-negative terms

Given a sequence $\{a_k\}_{k=0}^{\infty}$ such that $a_k \geq 0$, we say that $s_n = \sum_{k=0}^{n} a_k$ we say that s_n is a non-negative term series. Their main characteristic is that the partial sums are *non decreasing* in n:

$$s_n - s_{n-1} = a_n \geq 0 \Rightarrow s_n \geq s_{n-1} \, .$$

If a sequence is non-decreasing, we have a simple convergence criterion.

Theorem Given a non-decreasing sequence $\{s_n\}_{n=0}^{\infty}$ of real numbers we have that:

$$s_n < C \Rightarrow \exists \lim_{n\to\infty} s_n.$$

That is, if the partial sum of the series is *bounded*, then it is convergent.

Proof Using the fact that the sequence s_n is non-decreasing and bounded, all the terms are contained in the interval $I_0 = [s_0, C]$. We proceed now by a bisection argument. We divide I_0 in two halves, and we choose the only half part that contains an infinite number of terms (check why we have only one half with this property). We call this interval I_1, that has length $\frac{C-s_0}{2}$. We follow the same procedure with I_1, that is, we divide again in two halves and we keep the half part with infinite terms of the sequence. With this half we define I_2, with length $\frac{C-s_0}{2^2}$. Continuing in this way, we eventually obtain an intereval I_n of length $\frac{C-s_0}{2^n}$. Each os these intervals contains an infinite number of terms, leaving outside them a finite amount. This is a *Cauchy sequence* and it is therefore convergent.

Another proof: using that C is an upper bound of the numbers s_n, we can take the *minimum upper bound of this set* (that is, the *supremum*, see wikipedia) that is, if L is supremum of $\{s_n\}_{n=0}^{\infty}$, then $s_n \leq L$ for every n and moreover, for any other L' with the same property, we have that $L \leq L'$. We can see easily that L is the limit of the sequence s_n by the non-decreasing property. ■

If the sequence a_n alternate signs, then the partial sums do not satisfy the assumptions of the theorem. Consider $a_n = (-1)^n$. In this case:

$$s_n = \sum_{k=0}^{n} a_n = \begin{cases} 1 & \text{for even } n \\ 0 & \text{for odd } n \end{cases}$$

that is, s_n is bounded by $C = 1$, but is not convergent. The sequence s_n is not increasing nor decreasing.

Usually, integrals are used to bound the values of series by showing that each term corresponds to the integral over some interval. If the integral can be computed explicitly we obtain a bound. For example, if $a_n = \frac{1}{n^2}$, for $n \geq 1$, we can see that

$$\frac{1}{k^2} \leq \frac{1}{x^2} \quad \text{if } (k-1) \leq x \leq k\,.$$

Then we can write:

$$\frac{1}{k^2} \leq \int_{k-1}^{k} x^{-2}dx \Rightarrow s_n = 1 + \sum_{k=2}^{n} \frac{1}{k^2} \leq 1 + \int_{1}^{n} x^{-2}dx$$

The last integral can be computed explicitly, so we obtain a bound for s_n:

$$\int_{1}^{n} x^{-2}dx = -x^{-1}|_1^n = 1 - \frac{1}{n}\,.$$

That is, we can use the value 2 as a bound:

$$s_n \leq 1 + 1 - \frac{1}{n} < 2\,.$$

Using that the series is increasing, we conclude that it is convergent, because it is bounded.

Absolutely convergent series

From the partial sums $s_n = \sum_{k=0}^{n} a_k$ we can obtain another series with non-negative terms:

$$\hat{s}_n = \sum_{k=0}^{n} |a_k|\ .$$

Definition A series of real or complex numbers $\{s_n\}_{n=0}^{\infty}$ is *absolutely convergent* if the non-negative series$\{\hat{s}_n\}_{n=0}^{\infty}$ is convergent.

Using the previous property, we can say that a series is absolutely convergent if and only if $\hat{s}_n$ is bounded from above. On the other hand, if the series is absolutely convergent and $m < n$, then:

$$|s_n - s_m| = \left| \sum_{k=m+1}^{n} a_k \right| \leq \sum_{k=m+1}^{n} |a_k| = \hat{s}_n - \hat{s}_m \,.$$

Being $\hat{s}_n$ convergent, is is also a Cauchy series, and the last inequality shows that s_n is a Cauchy sequence. We have shown the following theorem:

Theorem Every absolutely convergent series is also convergent (the reciprocal is not true).

Alternating series

A series with positive and negative terms, both in infinite amount and distributed in an aleatory pattern are difficult to study. We consider a type of series that has a well defined pattern of positive and negative signs:

$$s_n = \sum_{k=0}^{n} (-1)^k a_k \,, \quad a_k \geq 0 \,.$$

The signs are alternating, and the series is called an *alternating series.*

We show the following result: if $a_{k+1} \leq a_k$ and also $\lim_{k\to\infty} a_k = 0$, then the alternating series is convergent. We start with $s_0 = a_0$, then $s_1 = a_0 - a_1 \geq 0$ because a_k is non-increasing, so all the following values are contained in the interval $I_1 = [s_0 - a_1, s_0]$. Continuing with $s_2 = a_0 - a_1 + a_2$ we see that the following values are inside $I_2 = [s_1, s_1 + a_2]$. We can say that all these intervals are "decreasing" in the following sense: $I_{n+1} \supset I_n$ for all n, and the length of I_n is exactly a_n. The intersection of the intervals is a point that is the limit of the series. Another way of showing this is that the terms indexed by even numbers form a *decreasing sequence,* while the terms with odd indices are an *increasing sequence.* Both sequences are positive and bounded (the first from below, the second from above) and then both of them are convergent. The difference between an even and and odd term is:

$$s_{2k+1} - s_{2k} = a_{2k+1} \to 0$$

and then they converge to the same limit. This result was proved by Leibniz (and is usually called *Leibniz criterion*).

Theorem Every alternating sequence such that the absolute value of its terms converges monotonically to zero is convergent.

Example The previous result shows that the partial sums:

$$s_n = \sum_{k=0}^{n} \frac{(-1)^k}{k}$$

is convergent. Notice that in this case, $\hat{s}_n$ is not convergent.

1.4. Series of functions

A function series is defined from a function sequence in the same way as a number series is defined from the sequence given by its general term:

$$s_n(x) = \sum_{k=0}^{n} f_k(x)$$

where x lies in an interval that may be the whole real line $\mathbb{R}$. This series of functions, denoted by $\sum f_n$, may converge pointwise, uniformly or diverge on some points, or in all of them. If the series is pointwise convergent to a function $S(x)$ we write:

$$S(x) = \lim_{n\to\infty} s_n(x) \ .$$

Criteria for uniform convergence

Cauchy Criterion: let f_n be a sequence of functions. The series $\sum f_n$ uniformly converges on a domain D if and only if for every $\varepsilon > 0$ there exists N_ε such that, for every pair n, m satisfying $n > m > N_\varepsilon$ we have that:

$$|f_{m+1}(x) + f_{m+2}(x) + \cdots + f_n(x)| < \varepsilon\,, \quad x \in D\,.$$

Weierstrass M test: Let $\{M_k\}_{k=0}^{\infty}$ be a sequence of positive numbers such that $\sum M_k$ is convergent, and moreover:

$$|f_k(x)| \leq M_k \quad \forall x \in I\,.$$

Then, the series $\sum f_k$ is *uniformly convergent* on D.

Weierstrass test is proved by applying Cauchy criterion as follows: given $\varepsilon > 0$ there exists N_ε such that

$$|s_n(x) - s_m(x)| \leq M_n + M_{n-1} + \cdots + M_{m+1} < \varepsilon \quad \forall x \in D$$

because $\sum M_k$ is convergent and satisfies itself the Cauchy criterion.

Convergence of derivatives and integrals

We can apply the uniform convergence and integration theorem to a series of functions. That is: if $s_n(x)$ uniformly converges to $S(x)$ in $[a, b]$, then the integrals of s_n on that interval, converge to the integrals of f:

$$\begin{aligned}\int_a^b S(x)\,dx &= \int_a^b \lim_{n\to\infty} s_n(x)\,dx = \lim_{n\to\infty}\int_a^b s_n(x)\,dx = \lim_{n\to\infty}\int_a^b \sum_{k=0}^{n} f_k(x)\,dx \\ &= \lim_{n\to\infty}\sum_{k=0}^{n}\int_a^b f_k(x)\,dx = \lim_{n\to\infty}\sum_{k=0}^{n} I_k\,.\end{aligned}$$

In other words, when the series is uniformly convergent, the integrals $I_k := \int_a^b f_k(x)\,dx$ are a series of numbers (not functions) whose partial sums converge to $I = \int_a^b S(x)\,dx$.
We can also apply immediately the convergence theorem for derivatives of function sequences. If $\sum f_n$ does converge pointwise to S on $[a,b]$, the derivatives f_n' exist and are continuous on the same interval, and $\sum f_n'$ uniformly converges to a function G, then S has a derivative such that $S'(x) = G(x)$, that is

$$S'(x) = \lim_{n\to\infty} \sum f_n'(x)\,.$$

Power series

One of the most common series of functions is the *power series*. It is defined by a general term of the form:

$$f_k(x) = a_k x^k\,, \quad a_k \in \mathbb{R}\,.$$

In this case, the partial sums are given by:

$$s_n(x) = \sum_{k=0}^{n} a_k x^k\,.$$

We can show the following result:

Theorem For every power series there exists a non-negative number R called *convergence radius*, such that:

- The series is absolutely convergent for $|x| < R$.
- The series is divergent for $|x| > R$.
- Convergence is not guaranteed for $|x| = R$.

The convergence radius can be determined by the formulae (when the limits involved exist):

$$R = \lim_{k\to\infty} \frac{a_k}{a_{k+1}} \qquad R = \lim_{k\to\infty} (a_k)^{1/k}\,,$$

that correspond to the *quotient* and *root* criteria respectively.

REMARK: if the convergence radius is zero, then the series is divergent for every x.

2. Fourier Series and Transforms

Bibliography: Partial Differential Equations of Mathematical Physics and Integral Equations, Ronald Guenther and John Lee. Prentice Hall 1988, Chapter 3.

2.1. Fourier series

Fourier series are series of trigonometric functions that arise in initial value problems in mathematical physics (for example, in the heat equation and the propagation of elastic waves) as well as in the mathematical theory of signals.

Let us assume that a function f is represented by means of a trigonometric series on the interval $-L \leq x \leq L$:

$$f(x) = A + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{\pi n x}{L}\right) + b_n \sin\left(\frac{\pi n x}{L}\right) \right) . \tag{2.1}$$

The coefficients A, a_n and b_n are determined by f, and the question is how can we find them. Some formal computations can give a clue about the adequate method. If the series is integrated term by term, using the following properties:

$$\int_{-L}^{L} \cos\left(\frac{\pi n x}{L}\right) dx = 0 \qquad \int_{-L}^{L} \sin\left(\frac{\pi n x}{L}\right) dx = 0 \,,$$

we would find that:

$$\begin{aligned} \int_{-L}^{L} f(x)\,\mathrm{d}x &= \int_{-L}^{L} A\mathrm{d}x + \int_{-L}^{L} \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{\pi n x}{L}\right) + b_n \sin\left(\frac{\pi n x}{L}\right) \right) \mathrm{d}x \\ &= A2L + \sum_{n=1}^{\infty} \left(a_n \underbrace{\int_{-L}^{L} \cos\left(\frac{\pi n x}{L}\right) \mathrm{d}x}_{=0} + b_n \underbrace{\int_{-L}^{L} \sin\left(\frac{\pi n x}{L}\right) \mathrm{d}x}_{=0} \right) \end{aligned}$$

and then:

$$A = \frac{1}{2L} \int_{-L}^{L} f(x)\,\mathrm{d}x \,.$$

Coefficients a_n y b_n can be found in a similar way by using the *orthogonality relationships*:

$$\int_{-L}^{L} \cos\left(\frac{\pi n x}{L}\right) \cos\left(\frac{\pi m x}{L}\right) dx = 0 \qquad m \neq n$$

$$\int_{-L}^{L} \sin\left(\frac{\pi n x}{L}\right) \sin\left(\frac{\pi m x}{L}\right) dx = 0 \qquad m \neq n$$

$$\int_{-L}^{L} \cos\left(\frac{\pi n x}{L}\right) \sin\left(\frac{\pi m x}{L}\right) dx = 0 \qquad \text{for every } m, n$$

and the identities:

$$\int_{-L}^{L} \sin^2\left(\frac{\pi n x}{L}\right) dx = \int_{-L}^{L} \cos^2\left(\frac{\pi n x}{L}\right) dx = L \qquad n \geq 1\,.$$

Then, by multiplying both sides of (2.1) by $\cos(m\pi x/L)$ and integrating term by term between $-L$ and L (using the orthogonality relationships and the previous identities) we find:

$$\int_{-L}^{L} f(x) \cos\left(\frac{m\pi x}{L}\right) dx = a_m L\,, \quad m \geq 1\,,$$

and solving for a_m:

$$a_m = \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{m\pi x}{L}\right) dx\,, \quad m \geq 1\,. \tag{2.2}$$

In the same way, multiplying by $\sin(m\pi x/L)$ we obtain:

$$b_m = \frac{1}{L} \int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) dx\,, \quad m \geq 1\,. \tag{2.3}$$

It is convenient to use the first of these formulae also for $m = 0$, to define A:

$$A = \frac{a_0}{2}\,, \tag{2.4}$$

because a_0 is given by:

$$a_0 = \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{0\pi x}{L}\right) dx = \frac{1}{L} \int_{-L}^{L} f(x)\, dx\,.$$

These formal computations suggest that, if a function admits an expansion in trigonometric series like (2.1), then A is given by (2.4) and the coefficients a_n ($n \geq 0$), b_n ($n \geq 1$) are given by (2.2-2.3).

Definition The trigonometric series, with the coefficients defined as in (2.2-2.3, 2.4), is called the *Fourier series* of f for the interval $[-L, L]$. The numbers a_n and b_n are called the *Fourier coefficients* of the function f.

In this context, one of the basic problems is to find conditions to be satisfied by f in order to have an equivalence between the sum of its Fourier series (that is, such that its Fourier series converges to the function). The formal arguments given before, indicate that (2.1) is valid for every continuous function but, unfortunately, there are continous functions such that their Fourier series do not converge to the original function.

We show some convergence criteria that guarantee the representation (2.1). We recall some definitions:

Definition A function f defined on $(-\infty, +\infty)$ is T-periodic if $f(x+T) = f(x)$ for all x and some $T > 0$. T is the *period* of the function if it is the smaller positive (ie. > 0) value with this property.

Definition A function f is *piecewise continuous* on $[a, b]$ if it is continuous except for a finite number of points in that interval, where it may have *jump discontinuities*. The function may not be well defined on that points. That means that, on that discontinuity points, the lateral limits exist and are finite, but they are not equal. If c is one of those points, we write:

$$\lim_{x \to c, x<c} f(x) = f(c^-) \qquad \lim_{x \to c, x>c} f(x) = f\left(c^+\right) .$$

Definition A function f is piecewise continuous over $\mathbb{R}$ if it is piecewise continuous on each finite interval $[a, b]$. In other words, f is continuous except for a finite amount of jump discontinuities on each finite interval.

Definition f is *piecewise smooth* on the interval $[a, b]$ if both f and f' (its derivative) are piecewise continuous on $[a, b]$. The derivative f' is not defined on the discontinuity points of f. f is piecewise smooth on $(-\infty, +\infty)$ if it is piecewise smooth over each finite interval.

We state now the following theorems:

Theorem Convergence. If the Fourier series of a periodic function f with period T, *uniformly converges* (to some limit) on one period, then the series converges to the function f.

Remark: The hypotheses do not say that the series converges to f, but merely that it is uniformly convergent to *some* function. The theorem states that in this case the Fourier series of f converges to f. Notice that the hypotheses imply that f must be continuous, just because it is a uniform limit of continuous functions.

One useful consequence is the following:

Corollary Let f be T periodic and continuous. If the series of absolute values of the coefficients $(\sum |a_n|, \sum |b_n|)$ are convergent, then the Fourier series of f does converge absolute and uniformly to f.

Theorem Dirichlet. Let f be T periodic and *piecewise smooth*. Then, the Fourier series of f is convergent for each $x \in (-\infty, +\infty)$ to

$$\frac{f(x^-) + f(x^+)}{2} ,$$

where $f(x^{\pm})$ is the lateral limit from each side. Moreover, the convergence is absolute and uniform on each closed subinterval containing no discontinuity points of the function f.

Corollary If f is piecewise smooth and continuous at the point x, its Fourier series evaluated on x converges to $f(x)$. If it is continous on all x, its Fourier series converges absolute and uniformly to f on $(-\infty, +\infty)$.

SOME PRACTICAL REMARKS: if the function is initially defined on an interval of the form $[0, L]$, it is always possible to extend it to the interval $[-L, L]$ *by an even* extension, that is, defining $f(x) := f(-x)$ for $x \in [-L, 0]$, or an *odd extension*, by means of $f(-x) := -f(x)$. Then it can be extended to the whole line $\mathbb{R}$ using a *periodic extension*, that is: $f(x + 2L) = f(x)$ (note: the period is $2L$) for $x \in [-L, L]$. After a periodic extension, we use the same letter f for the extended function, but whe must indicate explicitly if the function was extended in an even or odd manner. It must be kept in mind that for both cases we obtain different Fourier series, but if they converge, they have the same limit on $(0, L)$. When we study Fourier series it is convienient to assume that $L = \pi$, so $T = 2\pi$. If $L \neq \pi$ we can always change variables so that we have $L = \pi$ in the new variable. The Fourier series of f, for $L = \pi$ has the simpler expression:

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right) ,$$

with

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)\, dx , \quad n \geq 0 , \tag{2.5}$$

and

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)\, dx , \quad n \geq 1 . \tag{2.6}$$

2.2. Computing Fourier coefficients: examples

Complex form

Taking into account Euler's identity:

$$e^{ix} = \cos x + i \operatorname{sen} x$$

we are able to compute the coefficients in a compact form using one complex number instead of two real ones[1]:

$$a_n - i b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, e^{-inx} dx \tag{2.7}$$

This is useful when the integrand has a simpler expression in terms of complex numbers. A trivial example is the case $f(x) = e^x$ because we can integrate directly

[1] do not confuse this with the computation of the *complex Fourier coefficients,* that are defined later, when studying the Fourier transform. We will have that $c_n := \frac{a_n - i b_n}{2}$ for all n (we don't have to take apart the case $n = 0$ as in the real case).

without using integration by parts:

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} e^{x}e^{-inx}dx &= \frac{1}{\pi}\int_{-\pi}^{\pi} e^{(1-in)x}dx = \frac{1}{\pi}\left.\frac{e^{(1-in)x}}{1-in}\right|_{-\pi}^{\pi} \\ &= \frac{1}{\pi\left(1-in\right)}\left(e^{\pi}e^{-in\pi}-e^{-\pi}e^{+in\pi}\right) \\ &= \frac{(-1)^{n}}{\pi\left(1-in\right)}\left(e^{\pi}-e^{-\pi}\right)\end{aligned}$$

After this we have to compute the real and imaginary parts of the final expression, in order to find a_n and b_n. We do this easily by multiplying by $(1+in)$ above and below:

$$a_n - ib_n = \frac{(-1)^{n}\left(1+in\right)}{\pi\left(1+n^{2}\right)}\left(e^{\pi}-e^{-\pi}\right)$$

and we finally obtain:

$$a_n = (-1)^{n}\frac{\left(e^{\pi}-e^{-\pi}\right)}{\pi\left(1+n^{2}\right)} \quad b_n = -na_n\,.$$

Polynomials

To compute the Fourier series of a polynomial function

$$P\left(x\right) = c_m x^{m} + c_{m-1}x^{m-1} + \cdots + c_0$$

we can integrate by parts so that we reduce its degree one by one. For example (for $n \neq 0$):

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} P\left(x\right)\cos\left(nx\right)dx = \frac{1}{\pi}\left(P\left(x\right)\left.\frac{\sin\left(nx\right)}{n}\right|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} P'\left(x\right)\frac{\sin\left(nx\right)}{n}dx\right)$$

The expression $\frac{\sin(nx)}{n}$ when evaluated at $\pm\pi$ is equal to zero, so that we reduced by one the polynomial degree against which we must integrate:

$$a_n = -\frac{1}{n\pi}\int_{-\pi}^{\pi} P'\left(x\right)\sin\left(nx\right)dx\,,$$

so, in this way we continue reducing the polynomial degree, up to some point where we must integrate a trigonometric function. The calculation of the coefficient b_n is analogous. In these cases it is useful to separate even and odd powers, because we can take advantage of the integrals that cancel trivially. We can also use the complex compact expression, in this case (for $n \neq 0$):

$$a_n + ib_n = \frac{1}{\pi}\int_{-\pi}^{\pi} P\left(x\right)e^{inx}dx = \frac{1}{\pi}\left.P\left(x\right)\frac{e^{inx}}{in}\right|_{-\pi}^{\pi} - \frac{1}{in\pi}\int_{-\pi}^{\pi} P'\left(x\right)e^{inx}dx$$

In this case, $P\left(x\right)$ is a real function, so we compute only for $n \geq 0$ (it is not the same as in the case of a complex function where we must consider all the family $\left\{e^{inx}\right\}_{n\in\mathbb{Z}}$).

Powers of trigonometric functions

If $f(x) = (\cos x)^n$ or $f(x) = (\sin x)^n$ we can use trigonometric identities. On the other hand, we must first point out *what kind of expansion we are looking for*, especially if instead of $[-\pi, \pi]$ our interval is, for example, $[0, \pi]$ with an expansion in sines, or in cosines. Look at the following example:

$$f(x) = (\cos x)^2 \,, \qquad x \in [-\pi, \pi] \,.$$

It is an *even* function, that is $f(x) = f(-x)$, *so every coefficient* b_n is zero for that interval. The expansion we are looking for is as follows:

$$(\cos x)^2 = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx \,.$$

To find a_n we can use trigonometric identities, or the exponential complex expressions. Look both strategies:

a) **Trigonometric identities**. We write

$$(\cos x)^2 = \frac{1 + \cos 2x}{2} = \frac{1}{2} + \frac{1}{2}\cos 2x \,.$$

So, in the expansion we must have $a_0 = 1$ ($A = \frac{1}{2}$), $a_2 = \frac{1}{2}$, *and all the other coefficients are zero.*

b) **In terms of complex exponentials**:

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \Rightarrow (\cos x)^2 = \frac{1}{4}\left(e^{i2x} + 2 + e^{-i2x}\right)$$

So then:

$$(\cos x)^2 = \frac{1}{2} + \frac{1}{2}\left(\frac{e^{i2x} + e^{-i2x}}{2}\right) = \frac{1}{2} + \frac{1}{2}\cos 2x \,.$$

That is exactly the same as in a).

Trigonometric identities, the general case: Consider an even power of the sine or cosine, we can always write:

$$(\cos x)^{2k} = \left((\cos x)^2\right)^k = \left(\frac{1 + \cos 2x}{2}\right)^k \,,$$

then, expanding the binomial we will have powers of $\cos 2x$ (between 0 and k). The even powers are treated in the same way as before. If the power is odd, we can write it as a product of a unique cosine and another with an even power. Continuing in this way, we can use the following identity to reduce the exponent:

$$\cos(\alpha + \beta) + \cos(\alpha - \beta) = 2\cos\alpha\cos\beta$$

We can further reduce all the products and powers to sums of sines and cosines evaluated in integer multiples of x. When we have even powers of sines we use:

$$(\sin x)^{2k} = \left(\sin^2 x\right)^k = \left(\frac{1-\cos 2x}{2}\right)^k ,$$

and when we have odd powers we separate one of the factors and use:

$$\sin(\alpha+\beta) + \sin(\alpha-\beta) = 2\sin\alpha\cos\beta .$$

Example Odd power case, expansion of $(\sin x)^3$.

$$\begin{aligned}(\sin x)^3 &= \sin x\ (\sin x)^2 = \sin x\left(\frac{1-\cos 2x}{2}\right)\\ &= \frac{1}{2}\sin x - \frac{1}{2}\sin x\ \cos 2x\end{aligned}$$

now we have the product $\sin x\ \cos 2x$, and we use the trigonometric identity with $\alpha = x$, $\beta = 2x$:

$$2\sin x\ \cos 2x = \sin(3x) - \sin(x)$$

and gather all:

$$\begin{aligned}(\sin x)^3 &= \frac{1}{2}\sin x - \frac{1}{4}(\sin(3x) - \sin(x))\\ &= \frac{3}{4}\sin x - \frac{1}{4}\sin(3x) .\end{aligned}$$

Expansions on intervals different from $[-\pi, \pi]$

We must take into account which family of functions they are asking for. For example, if the interval is $[0, \pi]$, the orthogonal trigonometric families used are:

$$\{\cos nx\}_{n\geq 0} , \quad \{\text{sen}\, nx\}_{n\geq 1} .$$

These are two *different* families, that turn to be identical to the ones used on the interval $[-\pi, \pi]$ if we perform an odd or even extension (respectively) of the function to be expanded. For example, we may be asked to develop $(\cos x)^2$ with the cosine family, or with the sine family alone. As we are now on the interval $[0, \pi]$, the expansion on the sine family is not identically zero. We consider both cases.

The *even* extension of $(\cos x)^2$ from $[0, \pi]$ to $[-\pi, \pi]$ is identicall to the original function in the bigger interval. Then, its expansion in the cosine family on $[0, \pi]$ is the one already computed:

$$\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x .$$

However, if we consider the *odd* extension to $[-\pi, \pi]$, the new function *is different* from $(\cos x)^2$ on the big interval. When a function is odd on $[-\pi, \pi]$, we know that $a_n = 0$ for all n and, moreover:

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \underbrace{\underbrace{f(x)}_{\text{odd}}\underbrace{\sin(nx)}_{\text{odd}}}_{\text{even}}\, dx = \frac{2}{\pi}\int_0^{\pi} f(x)\sin(nx)\, dx\,.$$

We are looking for the expansion:

$$(\cos x)^2 = \sum_{n=1}^{\infty} b_n \sin(nx) \quad x \in [0, \pi]\ .$$

It is very important to take into account what kind of expansion we are looking for on $[0, \pi]$, because on $[-\pi, 0]$ this expansion will not be equal to the function $(\cos x)^2$ but to $-(\cos x)^2$. We now calculate the integrals:

$$\int_0^{\pi} (\cos x)^2 \sin nx\, dx$$

We write it in terms of complex exponentials (we can do it with trigonometric identities too):

$$\begin{aligned}\int_0^{\pi} \frac{1}{4}\left(e^{i2x} + 2 + e^{-i2x}\right)\left(\frac{e^{inx} - e^{-inx}}{2i}\right) dx &= \frac{1}{4}\int_0^{\pi}\left(\left(\frac{e^{i(n+2)x} - e^{-i(n+2)x}}{2i}\right)\right.\\ &\left. +2\left(\frac{e^{inx} - e^{-inx}}{2i}\right) + \left(\frac{e^{i(n-2)x} - e^{-i(n-2)x}}{2i}\right)\right) dx\\ &= \frac{1}{4}\int_0^{\pi} \left(\sin(n+2)x + 2\sin nx + \sin(n-2)x\right) dx\\ &\underbrace{=}_{n\neq 2} -\frac{1}{4}\left(\frac{\cos(n+2)x}{n+2} + 2\frac{\cos nx}{n} + \frac{\cos(n-2)x}{n-2}\right)\Bigg|_0^{\pi}\\ &= \frac{1}{4}\times\begin{cases} 0 & \text{if } n \text{ is even but not (yet) } 2\\ \frac{2}{n+2} + \frac{4}{n} + \frac{2}{n-2} & \text{if } n \text{ is odd}\end{cases}\end{aligned}$$

for $n = 2$ we have

$$\begin{aligned}\frac{1}{4}\int_0^{\pi} \left(\sin(n+2)x + 2\sin nx + \sin(n-2)x\right) \mathrm{d}x &= \frac{1}{4}\int_0^{\pi}\left(\sin 4x + 2\sin 2x\right) \mathrm{d}x\\ &= -\frac{1}{4}\left(\frac{\cos 4x}{4} + 2\frac{\cos 2x}{2}\right)\Bigg|_0^{\pi} = 0\,.\end{aligned}$$

Then, for all $n \geq 1$:

$$b_n = \frac{1}{\pi}\times\begin{cases} 0 & \text{for even } n\\ \frac{2}{n} + \frac{2n}{n^2-2} & \text{for odd } n\end{cases}$$

Notice that, on $[0, \pi]$, the same function $(\cos x)^2$ can be expanded in terms of cosines (with a finite expansion) or in terms of sines (with an infinite amount of terms). The choice of family will depend on the problem that we must solve.

2.3. Quadratic approximation

A basic practical problem in order to *effectively compute* a function is: Given a more or less complicated function f, find a simple function S, easy to compute, that approximates f as much as we want. The criteria used to determine wether S is a good approximation to f will depend on the practical problem at hand. For example, in some engineering or physics problems, the concept of energy is used to define the "distance" between functions. f and S will be close enough if the

$$\text{distance}(f, S) \equiv \left(\int_a^b |f(x) - S(x)|^2 \right)^{1/2}$$

is small, computed on on some interval $[a, b]$ defined by the practical problem. This is the so-called *quadratic approximation*, *least squares approximation* or L_2-approximation, and it plays a major role in applied mathematics and statistics. On the other hand, if our purpose is to provide a table of approximated values of f, we need that

$$|f(x) - S(x)|$$

be small enough for all value of x of the given interval. This concept is compatible with the notion of *uniform approximation* that we met before. In mathematics this is usually called *aproximation by the infinity norm, or* L_∞.

We will consider both "distances" for the case in which S is a trigonometric polynomial.

Definition A *trigonometric polynomial* if a function of the form

$$S_n(x) = \frac{1}{2}\alpha_0 + \sum_{k=1}^{n} (\alpha_k \cos(kx) + \beta_k \sin(kx)) \,, \tag{2.8}$$

where α_k, β_k are constants. S_n has *degree* n if $\alpha_n \neq 0$ or $\beta_n \neq 0$.

We want to solve the following problem: What is the best approximating trigonometric polynomial to a function f defined on $[-\pi, \pi]$ in the *least squares* sense? That is, we want to minimize the quadratic distance:

$$\int_{-\pi}^{\pi} (f(x) - S_n(x))^2 \, dx$$

by choosing adequately the coefficients of the trigonometric polynomial T_n. If we expand the integrand, we obtain a quadratic expression depending on the $2n + 1$ coefficients of the function S_n:

$$\int_{-\pi}^{\pi} (f(x) - S_n(x))^2 \, dx = \int_{-\pi}^{\pi} f(x)^2 \, dx - 2\int_{-\pi}^{\pi} f(x) S_n(x) \, dx + \int_{-\pi}^{\pi} S_n(x)^2 \, dx \,.$$

The first term is independent of the coefficients α_n, β_n, while the second and third terms certainly depend on them. We can check easily that:

$$\begin{aligned}\int_{-\pi}^{\pi} f(x) S_n(x)\, dx &= \int_{-\pi}^{\pi} f(x) \left(\frac{1}{2}\alpha_0 + \sum_{k=1}^{n} (\alpha_k \cos(kx) + \beta_k \sin(kx))\right) dx \\ &= \pi\left(\frac{1}{2}\alpha_0 a_0 + \sum_{k=1}^{n} (\alpha_k a_k + \beta_k b_k)\right).\end{aligned}$$

and also:

$$\begin{aligned}\int_{-\pi}^{\pi} S_n(x)^2\, dx &= \int_{-\pi}^{\pi} \left(\frac{1}{2}\alpha_0 + \sum_{k=1}^{n} (\alpha_k \cos(kx) + \beta_k \sin(kx))\right)^2 dx \\ &= \int_{-\pi}^{\pi} \left(\frac{1}{2}\alpha_0^2 + \sum_{k=1}^{n} \left(\alpha_k^2 \cos^2(kx) + \beta_k^2 \sin^2(kx)\right)\right) dx \\ &= \pi\left(\frac{\alpha_0^2}{2} + \sum_{k=1}^{n} \left(\alpha_k^2 + \beta_k^2\right)\right)\end{aligned}$$

Where we used the orthogonality relationships and integral identities for a_k and b_k. If we write all together:

$$\begin{aligned}\int_{-\pi}^{\pi} (f(x) - S_n(x))^2\, dx = \int_{-\pi}^{\pi} f(x)^2\, dx \\ + \pi\left(\frac{1}{2}\left(\alpha_0^2 - 2\alpha_0 a_0\right) + \sum_{k=1}^{n} \left(\left(\alpha_k^2 - 2\alpha_k a_k\right) + \left(\beta_k^2 - 2\beta_k b_k\right)\right)\right). \quad (2.9)\end{aligned}$$

We must compute the minimum of the quantity inside the parentheses, multiplying π. If we complete the squares we have that:

$$\begin{aligned}\frac{1}{2}\left(\alpha_0^2 - 2\alpha_0 a_0\right) + \sum_{k=1}^{n} \left(\alpha_k^2 - 2\alpha_k a_k\right) + \left(\beta_k^2 - 2\beta_k b_k\right) = \frac{1}{2}(\alpha_0 - a_0)^2 \\ + \sum_{k=1}^{n} (\alpha_k - a_k)^2 + (\beta_k - b_k)^2 - \left(\frac{a_0^2}{2} + \sum_{k=1}^{n} \left(a_k^2 + b_k^2\right)\right).\end{aligned}$$

The last term does not depend on α_k, β_k, so that, to minimize the quadratic expression we must have:

$$\alpha_k = a_k\,, k \geq 0 \qquad \beta_k = b_k\,, k \geq 1\,.$$

That is, the trigonometric polynomial that minimizes the quadratic distance is:

$$S_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \mathrm{sen} kx)$$

where a_k, b_k are given in (2.5-2.6). S_n is the partial sum of the Fourier series of f, up to degree n. We state this result as a tehorem.

Definition f is in $L_2[-\pi,\pi]$, or f is *square-integrable* on $[-\pi,\pi]$ means that $\int_{-\pi}^{\pi} f^2(x)\,dx < \infty$.

Theorem Let f be a square integrable function in $[-\pi,\pi]$. Then, the n-degree partial sum of its Fourier series provides the *best quadratic approximation* to f of all the trigonometric polynomials of degree n, that is:

$$\int_{-\pi}^{\pi} |f(x) - S_n(x)|^2 \leq \int_{-\pi}^{\pi} |f(x) - T_n(x)|^2$$

for all trigonometric polynomials T_n of degree *at most* n.

The hypothesis of the square integrability of f is the one that validates all the calculations done.

When we select $T_n = S_n$ in (2.9), we have that:

$$\int_{-\pi}^{\pi} (f(x) - S_n(x))^2\,dx = \int_{-\pi}^{\pi} f(x)^2\,dx - \pi\left(\frac{a_0^2}{2} + \sum_{k=1}^{n}\left(a_k^2 + b_k^2\right)\right), \qquad (2.10)$$

and, as the left hand side is a non-negative number:

$$\frac{a_0^2}{2} + \sum_{k=1}^{n}\left(a_k^2 + b_k^2\right) \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx\,,$$

for all $n \geq 1$. If we take limit for $n \to \infty$, we obtain the so-called *Bessel inequality*:

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}\left(a_k^2 + b_k^2\right) \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx\,.$$

Now we turn to uniform approximation by trigonometric polynomials. The main result here is known as the *Wierstrass approximation theorem*, and we omit the proof.

Theorem Weierstrass approximation Take $\varepsilon > 0$ and f a 2π-periodic and continuous function. Then, there exists a trigonometric polynomial S such that $|f(x) - S(x)| < \varepsilon$ for all $x \in \mathbb{R}$.

The main consequence of this theorem is the following:

Theorem If f is 2π-periodic and continuous, then its Fourier series converges to f in the least squares sense (or L_2 norm):

$$\lim_{n\to\infty}\int_{-\pi}^{\pi} |f(x) - S_n(x)|^2\,dx = 0\,,$$

where S_n the n-th partial sum of the Fourier series of f.

Proof Using Wierstrass Approximation Theorem, given $\varepsilon > 0$ (small as we want) we select a trigonometric polynomial S such that $|f(x) - S(x)| < \varepsilon$ for all x. Let n be the degree of S. We know that, among all the trigonometric polynomials

of degree n, Fourier's partial sum is optimal with respect to the quadratic norm, that is:

$$\int_{-\pi}^{\pi} |f(x) - S_n(x)|^2 \, dx \leq \int_{-\pi}^{\pi} |f(x) - S(x)|^2 \, dx \leq 2\pi\varepsilon^2 .$$

As ε can be arbitrarily small (then n can be arbitrarily large) and taking into account that:

$$\int_{-\pi}^{\pi} |f(x) - S_m(x)|^2 \, dx \leq \int_{-\pi}^{\pi} |f(x) - S_n(x)|^2 \, dx \quad \text{if } m > n ,$$

(see (2.10)), then the limit is be zero■

Corollary PARSEVAL IDENTITY If f is 2π-periodic and continuous, then

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2 \, dx = \frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}\left(a_k^2 + b_k^2\right) ,$$

where a_k, are the Fourier coefficients of f.

The last relationship is analogous to *Pythagoras Theorem* of Euclidean geometry (the left hand side is like the length of the hypotenuse of right triangle). An immediate consequence of Parseval's identity is the following important result:

Theorem If the Fourier coefficients of a continuous function on $[-\pi, \pi]$ are all equal to zero, then the function is identically zero. Therefore, if two functions (continous and 2π-periodic) have the same Fourier coefficients, they must be identical.

Proof If all the Fourier coefficients of f are zero, from Parseval's identity we obtain:

$$\int_{-\pi}^{\pi} f(x)^2 \, dx = 0 .$$

On the other hand, f^2 is a continuous and non-negative function, with integral equal to zero, it must be the zero function (if for some x_0 we had $f(x_0) \neq 0$, then there is a small interval around x_0, say $[x_0 - \delta, x_0 + \delta]$ such that $\int_{x_0-\delta}^{x_0+\delta} f(x)^2 \, \mathrm{d}x > 0$,and then $\int_{-\pi}^{\pi} f(x)^2 \, \mathrm{d}x$ cannot be zero). Moreover, if f and g have the same coefficients, then $f - g$ have its coefficients all zero, and we conclude (by the previous part) that $f - g = 0$, then $f = g$ ■

Proof of the Convergence Theorem

We can now prove the Convergence Theorem of the previous section. To this end, we will apply some of the convergence theorems that we saw before. We choose $L = \pi$ to make the expressions simpler. We have f, 2π-periodic and continuous function, and we assume that its Fourier series is uniformly convergent (to some function). Let $\hat{f}$ be the limit of this series. A previous result asserts that $\hat{f}$ is also a continuous function, because it is the uniform limit of continuous functions (the partial sums of a Fourier series is a continuous function). Moreover, the Fourier coefficients of $\hat{f}$ can be computed as limits of the partial sums of the Fourier series (by uniform convergence). This implies that the coefficients of $f - \hat{f}$ (continuous) are all zero. Then $f - \hat{f} = 0$, that is, *f is the uniform limit of its own Fourier series.*

2.4. Fourier transform

Fourier expansions are applied to periodic functions. This limitation is not serious, because any function can be extended to the whole line in a periodic way. But a non-periodic function (defined in the whole $\mathbb{R}$) cannot be represented by a Fourier series. Therefore, this method must be restricted to functions defined on finite intervals. To apply a similar representation in the case of an infinite interval we must use the so-called *Fourier transform.* We will motivate its definition omiting detailed proofs.

We start by showing how a Fouirer series expansion converges to an integral representation when the length of the interval grows to infinity. Consider a continuous function f defined in the whole real line. We denote its *constraint* (or projection) to the interval $[-L, L)$ by f_L, that is:

$$f_L(x) = f(x) \quad -L \leq x < L\,.$$

We extend now f_L *periodically* to the whole line. We can see easily that:

$$f(x) = \lim_{L\to\infty} f_L(x)\ .$$

The function thus extended may be discontinuous on the interval boundaries. But, if the original function f is piecewise smooth, Dirichlet's theorem (the second convergence theorem) guarantees that the Fourier series of f_L converges uniformly and absolutely on any closed subinterval with no discontinuity points. It is enough to consider a smaller interval, say $[-L+a, L-a]$ with $a > 0$ small, and in that interval we can ensure convergence:

$$f_L(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n(L)\cos\frac{n\pi x}{L} + b_n(L)\,\mathrm{sen}\frac{n\pi x}{L}\right) \tag{2.11}$$

where

$$a_n(L) = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)\mathrm{d}x$$

and

$$b_n(L) = \frac{1}{L}\int_{-L}^{L} f(x)\,\mathrm{sen}\left(\frac{n\pi x}{L}\right)\mathrm{d}x\,.$$

Notice that we guaranteed the convergence in a smaller interval, but we refer to the series in the whole interval $[-L, L]$. Our plan is to consider the limit for $L \to \infty$ in (2.11). In this case, we must ensure the existence of the integrals, so we assume that:

$$\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x < \infty\,.$$

For our purpose, it is easier to deal directly with the complex expansion of the function:

$$f_L(x) = \sum_{-\infty}^{\infty} c_n e^{in\pi x/L}$$

where:

$$c_n = \frac{1}{2L}\int_{-L}^{L} f(s)\, e^{-in\pi s/L} \mathrm{d}s\,.$$

If we consider both expressions together (representation + coefficients) we obtain:

$$f_L(x) = \sum_{-\infty}^{\infty}\left(\frac{1}{2L}\int_{-L}^{L} f(s)\, e^{-in\pi s/L}\mathrm{d}s\right) e^{in\pi x/L}\,.$$

Now consider the infinite sum as an aproximation of an integral in $(-\infty,\infty)$ for big L. We take $\xi_n := \frac{n\pi}{L}$, so $\Delta\xi_n = \frac{\pi}{L} \approx \mathrm{d}\xi$ gives:

$$\begin{aligned} f_L(x) &= \frac{1}{2\pi}\sum_{-\infty}^{\infty}\left(\frac{\pi}{L}\int_{-L}^{L} f(s)\, e^{-i\frac{n\pi}{L}s}\mathrm{d}s\right) e^{i\frac{n\pi}{L}x} \\ &= \frac{1}{2\pi}\sum_{-\infty}^{\infty}\left(\int_{-L}^{L} f(s)\, e^{-i\xi_n s}\mathrm{d}s\right) e^{i\xi_n x}\Delta\xi_n \\ &\underset{\text{for large } L}{\approx} \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(s)\, e^{-i\xi s}\mathrm{d}s\right) e^{i\xi x}\mathrm{d}\xi \end{aligned}$$

The function defined in the interior integral is called the *Fourier transform* of the function f:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(s)\, e^{-i\xi s}\mathrm{d}s\,, \tag{2.12}$$

that is finite, because f is absolutely integrable. The last equality tells us that we can recover the value of f using what is called the *inverse Fourier* transform:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{i\xi x}\,\mathrm{d}\xi\,,$$

where the last integral is taken in the sense of the *Cauchy principal value* that is, taking limits simetrically around zero ($\int_{-\infty}^{+\infty} \equiv \lim_{L\to\infty}\int_{-L}^{+L}$). We summarize the discussion in the following theorem:

Theorem Fourier Transform Let f be continuous, piecewise smooth and absolutely integrable, then we have the following integral representation:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{i\xi x}\mathrm{d}\xi\,, \tag{2.13}$$

where $\hat{f}$ is given in (2.12).

Now we turn to the relationship between the Fourier transform of f and the derivatives of f. We assume that both f and f' are continuous and *absolutely integrable* on $(-\infty,\infty)$. Integrating by parts (2.12), we obtain a relationship between $\hat{f}$ and $\hat{f}'$:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(s)\, e^{-i\xi s}\mathrm{d}s = \lim_{L\to\infty}\left(-\frac{e^{-i\xi s}}{i\xi} f(s)\Big|_{-L}^{L} + \int_{-L}^{L}\frac{e^{-i\xi s}}{i\xi} f'(s)\,\mathrm{d}s\right)$$

The fundamental theorem of calculus tells us that:

$$f(x) = f(0) + \int_0^x f'(s)\,\mathrm{d}s\,.$$

As f' is integrable, the limit of $f(x)$ for $x \to \infty$ exists, and then it must be zero (because f itself is absolutely integrable). Using this we can write:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} \frac{e^{-i\xi x}}{i\xi} f'(s)\,\mathrm{d}\xi \Rightarrow \hat{f}'(\xi) = i\xi \hat{f}(\xi)\ .$$

That means: there is no need to compute further integrals in order to find the transform of the derivatives of f: it is enough to multiply by $-i\xi$. If we apply this result repeatedly we prove the following theorem:

Theorem Given $f,\ f',\ \dots f^{(n)}$ continuous and absolutely integrable in $(-\infty, \infty)$. Then, $f,\ f',\ \dots f^{(n)}$ have Fourier transforms given by:

$$\widehat{\left(f^{(k)}\right)}(\xi) = (i\xi)^k\, \hat{f}(\xi) \quad k = 0, \dots, n\,. \tag{2.14}$$

2.5. Application to signal filtering

We will see an application of the Fourier transform in signal processing. If we consider f as a *signal* definided on the time domain t, its transform $\hat{f}$ is defined on the *frequency domain* ω. Our main objective is to describe the so-called *frequency filter*. In intuitive terms it means that, given a signal f, once we transform it into the frequency domain via Fourier transform, we *reduce* the frequency interval to the desired one and then reconstruct a *filtered signal* $f_s(t)$ by means of the inverse transform.

We describe now the filtering process in mathematical terms. Consider a function χ_I that is equal to the value 1 on an interval I and is zero out of this interval:

$$\chi_I(\omega) = \begin{cases} 1 & \omega \in I\,, \\ 0 & \omega \notin I\,. \end{cases}$$

Given a signal $\hat{f}(\omega) = \mathcal{F}[f](\omega)$ on the frequency domain (that is, by Fourier transforming the signal $f(t)$), we are interested in finding the function $f_I(t)$ such that

$$\hat{f}_I(\omega) = \hat{f}(\omega)\,\chi_I(\omega)\ .$$

in other words, we are looking for the signal f_I such that its Fourier transform is the product of the transform of the original signal and the function χ_I. For this purpose, we will state the following theorem, called the *convolution theorem*. We fist provide the following definition:

Definition Given $f,\, g : \mathbb{R} \to \mathbb{R}$, the *convolution* of f and g, denoted by $f * g$ is the function defined by the following integral:

$$f * g\,(x) = \int_{-\infty}^{\infty} f\,(x-y)\, g\,(y)\, \mathrm{d}y \tag{2.15}$$

provided that the improper integral exists.

Theorem CONVOLUTION Let $f,\, g$ be two absolutely integrable functions on $\mathbb{R}$. Then $f * g$ exists, is absolutely integrable and moreover, for every $\omega \in \mathbb{R}$:

$$\begin{aligned} \widehat{f * g}\,(\omega) &= \hat{f}\,(\omega) \cdot \hat{g}\,(\omega) \\ \widehat{f \cdot g}\,(\omega) &= \hat{f} * \hat{g}\,(\omega) \end{aligned}$$

Proof (part of it) The proof depends on the so-called *Fubini's Theorem* that, under certain conditions, allows to interchange the order of integration. We will assume that $f * g$ exists (that is, that the integral is well defined) and that Fubini's theorem is applicable. Then we will prove the first equality, that is the one needed for the frequency filter.

$$\begin{aligned} \widehat{f * g}\,(\omega) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f\,(x-y)\, g\,(y)\, dy \right) e^{-i\omega x} \mathrm{d}x \\ &= \int_{-\infty}^{\infty} g\,(y) \left(\int_{-\infty}^{\infty} f\,(x-y)\, dx \right) e^{-i\omega(x-y+y)} \mathrm{d}y \\ &= \int_{-\infty}^{\infty} g\,(y) \left(\int_{-\infty}^{\infty} f\,(x-y)\, e^{-i\omega(x-y)} dx \right) e^{-i\omega y} \mathrm{d}y \\ &\underset{u=x-y}{=} \int_{-\infty}^{\infty} g\,(y) \left(\int_{-\infty}^{\infty} f\,(u)\, e^{-i\omega u} du \right) e^{-i\omega y} \mathrm{d}y \\ &= \int_{-\infty}^{\infty} g\,(y)\, \hat{f}\,(\omega)\, e^{-i\omega y} dy = \hat{f}\,(\omega) \cdot \hat{g}\,(\omega) \end{aligned}$$

■

We see now how to apply the convolution theorem to construct a so-called *low pass filter*. This is the case when we limit the frequency to the interval $[-\delta, \delta]$ for $\delta > 0$. Consider a signal $f\,(t)$ and compute its Fourier transform $\hat{f}\,(\omega)$. Then we *filter* it to the given interval by taking the product $\hat{f}\,(\omega) \cdot \chi_{[-\delta,\delta]}\,(\omega)$. If we found the function $g\,(t)$ such that:

$$\hat{g}\,(\omega) = \chi_{[-\delta,\delta]}\,(\omega)\ ,$$

the convolution theorem allows us to compute:

$$\widehat{f * g}\,(\omega) = \hat{f}\,(\omega) \cdot \hat{g}\,(\omega) = \hat{f}\,(\omega) \cdot \chi_{[-\delta,\delta]}\,(\omega)\ ,$$

so, the filtered signal is:

$$f_\delta\,(t) = f * g\,(t)\ .$$

There are two steps to find f_δ:
1) Compute the inverse transform $g(t)$ of $\chi_{[-\delta,\delta]}(\omega)$, by using (2.13):

$$\begin{aligned} g(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\chi_{[-\delta,\delta]}(\omega)\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\delta}^{\delta} e^{i\omega t}d\omega = \frac{1}{2\pi}\left.\frac{e^{i\omega t}}{it}\right|_{-\delta}^{\delta} \\ &= \frac{1}{2\pi}\frac{e^{i\delta t}-e^{-i\delta t}}{it} = \frac{\sin(\delta t)}{\pi t}\,. \end{aligned}$$

2) Compute the convolution of f and g.

$$f_\delta(t) = (f*g)(t) = \int_{-\infty}^{\infty} f(t-s)\,\frac{\sin(\delta s)}{\pi s}\mathrm{d}s\,.$$

3. First Order Differential Equations

Bibliography: A first course in differential equations, Frank G. Hagin. Prentice Hall 1975. Chapters 1 and 2.

3.1. Introduction

The mathematical representation of models usually involves rates of quantities describing prescribed laws. For example, one of the oldest models for population growth (Malthus, 1798) is represented by the equation:

$$\frac{\mathrm{d}P}{\mathrm{d}t}(t) = kP(t) \ .$$

This equation implies that the rate of growth of a population P at time t is proportional to the amount of population, and the factor k. Sometimes, the independent variable (in this case, t) is omitted, when it can be guessed from the context. The equation involves a derivative of the function that we want to find, so this type of equation is called *differential equation*.

The mathematical models for population growth are only an approximation to the real case. This happens because we do not know in detail all the processes involved in the complex whole system. The Malthus model, for example, takes into account population growth under environmental invariance. When environment saturation is considered, the equation is a bit more complicated (Verhulst, 1836):

$$\frac{\mathrm{d}P}{\mathrm{d}t}(t) = kP(t)\left(1 - \frac{P(t)}{K}\right) \ .$$

Where K is a saturating value for the population (note that when $P = K$, then the rate of growth is zero).

Both equations can also be found in different disciplines (economy, chemistry, etc.), turning differential equations into a kind of *universal* language for describing models. Variables may have different interpretations, but the basic laws described are identical. In other words, the same equation can describe similar mechanisms in different contexts. That's why they are important and that's why we study them.

The fundamental equation of motion in classical mechanics (Newton's second law) is a differential equation:

$$mx''(t) = F \ .$$

This tells us that the position x of a particle with mass m at time t must be a function of time such that its second derivative times the mass must be equal to the applied force. The nature of this force depends on the physical context (friction, electricity, gravitation), but the law is *universal*, that is, it can be applied to any kind of force, independently of its nature. In this case, the unknown function has two derivatives, so the equation is of *second order.* We want to find a function $x(t)$ that *solves the equation*; such a function is known as a *solution* of the differential equation. The equation may be very difficult to solve if F depends explicictly on x. Take into account, that F may depend on x' too (such case arises when dealing with *friction forces*). This dependence gives rise to different types of equations.

Since 1950 computer technology improved at a very rapid pace, increasing the computation velocity and the storage capacity. These improvements allow to solve modeling problems by means of *numerical methods* that formerly where impossible to implement due to the time cost involved.

Some definitions

A differential equation involves one or more derivatives of an unknown function. This function may have one or more independent variables. If the equation involves derivatives taken with respect to different variables, the equation is called *partial derivatives equation* (or PDE). An example of such an equation is the *heat equation*:

$$\frac{\partial u}{\partial t}(x,t) = \kappa \frac{\partial^2 u}{\partial x^2}(x,t) \ .$$

An equation involving derivatives with respect to only one of its variables, such as $P'(t) = kP(t)$, is called *Ordinary Differential Equation* (ODE). The *order* of a differential equation is the order of the highest derivative of the unknown function in the equation. We refer to a *solution* when we have a function that, with all its derivatives, satisfies the equation over some interval for the independent variables.

For example, the equation:

$$x'(t) - 2x(t) = 0$$

has a *solution* $x(t) = e^{2t}$ for $-\infty < t < \infty$. But it admits as a solution any of the functions of the form $x(t) = Ce^{2t}$ for $C \in \mathbb{R}$. That is, the equation may have an *infinite number of solutions.* So, part of the problem is how to find the so-called *general solution* of the equation, that is, the complete set of solutions. From this set we must extract a *particular solution* when we impose other conditions (such as initial or boundary conditions).

Another elementary example is given by the differential equation

$$y''(t) = 2\,.$$

What is the general solution of this equation? This problem can be approached in the following way: if two functions f, g have identical derivatives, that is $f'(x) = g'(x)$ on some domain, then, $f(x) = g(x) + A$ where A is a constant. We know that the

derivative of the function $f(t) = 2t$ is the constant 2, then we must have (note that one derivative dissapears):

$$y'(t) = 2t + A$$

We also know that the derivative of $t^2 + At$ is equal to $2t + A$, then, applying once more the theorem we have that:

$$y(t) = t^2 + At + B\,.$$

This suggest that, if the differential equation is of order two, we must have *two arbitrary constants* in the general solution. This is an important property that we will study later for the so-called linear equations.

Ordinary differential equations of the first order are usually asociated to *initial conditions*. In the case of an equation that rules the position of a particle as a function of time, this is equivalent to stating an initial position. We will not be able to find the actual position if we do not provide a starting point. In general terms, a first order initial value problem may be written as an implicit equation for $x(t)$ and its derivatives:

$$F\left(t, x, x'\right) = 0 \quad x(t_0) = x_0\,.$$

If the first derivative can be "cleared" from the equation this equation can be written in the form:

$$x' = f(t, x) \quad x(t_0) = x_0\,.$$

Under regularity conditions over f, it can be shown that there exists one and *only one* solution to this problem. This result does not help if we want to find concrete solutions, but it is useful to know *existence* and *uniqueness* of solutions if we want to compute approximate solutions.

Care must be taken when we try to integrate the differential equation directly. For example, using the Fundamental Theorem of Calculus we can find solutions to the following initial value problem:

$$x'(t) = t\,, \quad x(t_0) = x_0\,.$$

In this case, it is enough to integrate both sides between t_0 and t:

$$\int_{t_0}^{t} x'(s)\,ds = \int_{t_0}^{t} s ds \Rightarrow x(t) - x(t_0) = \frac{t^2}{2} - \frac{t_0^2}{2}\,.$$

By using the initial condition we obtain:

$$x(t) = \left(x_0 - \frac{t_0^2}{2}\right) + \frac{t^2}{2}\,.$$

This problem is solved by merely calculating a definite integral. Now consider the case:

$$x'(t) = x(t) + t\,, \quad x(t_0) = x_0\,.$$

Here, if we integrate directly, we obtain an *integral equation* for $x(t)$:

$$x(t) - x_0 = \int_{t_0}^{t} x(s)\, ds + \frac{t^2}{2} - \frac{t_0^2}{2}\,.$$

That is, we did not find an explicit solution, we merely transformed one type of equation into another.

3.2. Variable separation

In this section we will use the letter y to denote the unknown function and x for the independent variable. We will look for a function $y(x)$ such that its derivative $y' = \frac{\mathrm{d}y}{\mathrm{d}x}$ satisfies a first order equation of the form:

$$y' = f(x, y)\,. \tag{3.1}$$

We will describe a method of solution for a special class of equations called *separable equations*. When the function f in (3.1) has the form:

$$f(x, y) = p(x)\, q(y)$$

this is known as a *separable equation*, and the technique of integration is called *separation of variables*. In other words, the method is applicable when the right hand side can be factorized with two functions of only one variable each (respectively, x and y).

Let us call

$$g(y) := \frac{1}{q(y)}\,,$$

Then we can write the equation as follows:

$$y' = \frac{p(x)}{g(y)} \Rightarrow g(y)\, y' = p(x)\,.$$

Now, let us assume that g has a primitive function G with respect to the y variable, that is

$$\frac{\mathrm{d}G}{\mathrm{d}y}(y) = g(y)\,,$$

then, using the chain rule, we can write:

$$\frac{\mathrm{d}\left[G(y(x))\right]}{\mathrm{d}x} = \frac{\mathrm{d}G}{\mathrm{d}y}(y)\,\frac{\mathrm{d}y}{\mathrm{d}x} = g(y)\, y'\,.$$

So, we have that:

$$\frac{\mathrm{d}\left[G(y(x))\right]}{\mathrm{d}x} = p(x)$$

If P is a primitive of p, we have:

$$G\left(y\left(x\right)\right) = P\left(x\right) + C$$

where C is aconstant. In some cases (we will see some examples), the function y can be cleared from this implicit expression, and the constant C can be find using initial conditions.

So, in this case the problem is reduced to finding *two primitive functions*, one for the function g, with respect to the variable y, and one for the function p, with respect to x. The final equation can be solved for y if G is a standard function.

Example Solve the following differential equation for the given initial value:

$$y' = x^3 e^{-2y}\,, \qquad y\left(1\right) = 0\,.$$

In this case, $g\left(y\right) = e^{2y}$ and $p\left(x\right) = x^3$. We compute both primitives:

$$G\left(y\right) = \int e^{2y} dy = \frac{e^{2y}}{2} + C_1 \qquad P\left(x\right) = \int x^3 dx = \frac{x^4}{4} + C_2\,.$$

Then, the general solution of our problem can be written as:

$$\frac{e^{2y}}{2} = \frac{x^4}{4} + C\,,$$

for some constant $C = C_2 - C_1$. Using now that for $x_0 = 1$ we must have $y_0 = 0$, we have that:

$$\frac{e^{2y_0}}{2} = \frac{x_0^4}{4} + C \Rightarrow C = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}\,.$$

we conclude that

$$\frac{e^{2y}}{2} = \frac{x^4}{4} + \frac{1}{4}\,.$$

In this particular case we can solve for y:

$$y\left(x\right) = \frac{1}{2}\log\left(\frac{1+x^4}{2}\right)\,.$$

We can check that for $x = 1$, y attains the value 0. Now, we check that the function satisfies the differential equation. We find first the derivative of y:

$$y' = \frac{2x^3}{1+x^4}$$

On the other hand:

$$x^3 e^{-2y} = x^3 e^{-\log\left(\frac{1+x^4}{2}\right)} = \frac{2x^3}{1+x^4}\,.$$

So, y is a solution of the initial value problem ■

3.3. The first order linear equation

The differential equation is called *linear* when the function f on the right hand side it is linear in the y variable, that is:

$$f(x,y) = p(x)\,y + q(x)\ . \tag{3.2}$$

If we consider it as a function of the variable y we have a straight line with slope $p(x)$ and ordinate $q(x)$.

Definition When the function $q(x)$ is zero we have a linear *homogeneous* equation of the first order.

Definition A solution to (3.2) that has no *undetermined parameters* or free constants is a *particular solution.* If, on the other hand, our solution contains free parameters that describe *the whole set of solutions,* we have a *general solution.*

We will describe the relationship between the solution of the homogeneous equation, the particular solution and the general solution. Assume that $y_p(x)$ is a particular solution that we found by some method, and has no free parameters. If we denote by $y(x)$ the general solution, we have that both functions satisfy the same equation:

$$\begin{aligned} y'(x) &= p(x)\,y(x) + q(x) \\ y_p'(x) &= p(x)\,y_p(x) + q(x) \end{aligned}$$

By taking the difference between both equations we have that:

$$(y - y_p)' = p(x)\,(y - y_p)$$

that is, the function

$$y_h = y - y_p$$

satisfies the homogeneous equation. In other words, if we know a particular solution to our problem (ie, without free parameters) and we are able to calculate the general solution to the homogeneous equation (that is easier than the original equation) we can calculate the general solution to the original equation by writing:

$$y = y_h + y_p\,.$$

We have the following result

Theorem *The general solution to a linear non-homogeneous equation (3.2) is always the sum of the general solution to the homogeneous equation and a particular solution.*

Constant coefficients

The easiest cases to solve ar the ones wher p and q are *constants.* For example, if $p \equiv 0$ and $q(x) \equiv q$ is a constant, we have by direct integration (y is not on the right hand side):

$$y' = q \Rightarrow y(x) = qx + C$$

where C is an arbitrary integration constant.

Let us assume now that $p(x) \equiv p$ is constant and $q \equiv 0$. This is identical to the Malthus model that we showed before. We have now a *separable equation*:

$$y'(x) = py(x) \ .$$

If $y(x) \neq 0$ in the working interval, we can write:

$$\frac{y'(x)}{y(x)} = p \Rightarrow (\log |y(x)|)' = p \, .$$

We are now in the former case and we can integrate directly:

$$\log(|y(x)|) = px + C \, ,$$

where C is an arbitrary constant. Clearing $|y|$, we find:

$$|y(x)| = e^{px+C} \Rightarrow |y(x)| = Ke^{px} \qquad K := e^C \, .$$

Removing the module bar we have:

$$y(x) = \pm Ke^{px} \qquad K > 0 \, ,$$

so we can say that:

$$y(x) = \tilde{K}e^{px} \qquad \tilde{K} \in \mathbb{R} \, ,$$

that is, $\tilde{K}$ is an arbitrary real constant (positive or negative). To find it we consider the initial data $y(x_0) = y_0$. In this case:

$$y_0 = \tilde{K}e^{px_0} \Rightarrow \tilde{K} = y_0 e^{-px_0}$$

and plugin this value in the solution we obtain:

$$y(x) = y_0 e^{p(x-x_0)} \, .$$

The main idea to integrate any linear first order equation is to transform it in such a way that all the "y's" are put together under one derivative. Let us try to integrate (p and q are constants):

$$y' = py + q,$$

writing it as

$$y' - py = q \, .$$

We cannot divide by y, because it will appear on the right hand side. But we can put together the terms involving y under the same derivative if we multiply both sides by e^{-px}:

$$e^{-px}y' - e^{-px}py = e^{-px}y' + \left(e^{-px}\right)' y = \left(e^{-px}y\right)' \, .$$

We can write now the equation as:

$$\left(e^{-px}y\right)' = q\,, \tag{3.3}$$

and then we can integrate it directly:

$$e^{-px}y\left(x\right) = qx + C$$

Clearing for $y\left(x\right)$,we find:

$$y\left(x\right) = e^{px}\left(qx + C\right)\,.$$

By imposing an initial condition we can also do this by means of a definite integral of (3.3) between x_0 and x:

$$e^{-px}y\left(x\right) - e^{-px_0}y_0 = q\left(x - x_0\right) \Rightarrow y\left(x\right) = e^{p(x-x_0)}y_0 + e^{px}q\left(x - x_0\right)\,.$$

Variable coefficients

The case in which p and q depend on x (that is, they are not constants anymore) can be solved analogously. The key idea is to put together the terms involving the function y in only one derivative. We move the term with y to the left hand side::

$$y' - p\left(x\right)y = q\left(x\right)\,.$$

We must look for an *integrating factor*, that we call $m\left(x\right)$. Multiplying both sides by $m\left(x\right)$ (to be found) we have that:

$$m\left(x\right)y' - m\left(x\right)p\left(x\right)y = m\left(x\right)q\left(x\right)\,, \tag{3.4}$$

if we want to write the left hand side as a unique derivative, m must satisfy:

$$\left(m\left(x\right)y\right)' = m\left(x\right)y' + m'\left(x\right)y\,, \tag{3.5}$$

that is, when $m\left(x\right)$ satisfies the equation:

$$m'\left(x\right) = -m\left(x\right)p\left(x\right)\,, \tag{3.6}$$

we may write, by equating the left hand side of (3.4) with the right hand side of (3.5):

$$\left(m\left(x\right)y\right)' = m\left(x\right)q\left(x\right)\,, \tag{3.7}$$

and we can integrate directly between x_0 and x:

$$m\left(x\right)y\left(x\right) - m\left(x_0\right)y\left(x_0\right) = \int_{x_0}^{x} m\left(s\right)q\left(s\right)ds$$

where we can solve for y (if $m\left(x\right) \neq 0$):

$$y\left(x\right) = \frac{m\left(x_0\right)}{m\left(x\right)}y\left(x_0\right) + \frac{1}{m\left(x\right)}\int_{x_0}^{x} m\left(s\right)q\left(s\right)ds\,.$$

The only remaining question is if we are able to calculate a function m such that $m'(x) = -m(x)\,p(x)$. We must integrate the equation:

$$\frac{m'(x)}{m(x)} = -p(x) \Rightarrow (\log|m|)' = -p(x)\ .$$

So we have that:

$$\log|m(x)| = -P(x)$$

where $P(x) := \int^x p(s)\,\mathrm{d}s$ is a *primitive* of p. Then we solve for $m(x)$ (any m like this works, so we take a positive one):

$$m(x) = e^{-P(x)}\,.$$

Notice that $P(x)$ is defined up to an additive constant; nevertheless, we need only one function $m(x)$ satisfying the condition of an integrating factor, there is no need to have a whole family of functions. If we replace now this value of m in the expression for the complete solution we have that:

$$\begin{aligned} y(x) &= e^{P(x)-P(x_0)}y(x_0) + e^{P(x)}\int_{x_0}^{x} e^{-P(s)}q(s)\,\mathrm{d}s \\ &= e^{P(x)-P(x_0)}y_0 + \int_{x_0}^{x} e^{P(x)-P(s)}q(s)\,\mathrm{d}s \end{aligned}$$

Example Consider the differential equation

$$y' = xy + 1$$

with initial condition

$$y(0) = 2\,.$$

In this case we have that $p(x) = x$ and $q(x) = 1$, with $x_0 = 0$, $y_0 = 2$. We calculate first a primitive of p:

$$P(x) = \frac{x^2}{2}\,.$$

Then, the solution is given by:

$$y(x) = 2e^{\frac{x^2}{2}} + e^{\frac{x^2}{2}}\int_0^x e^{-\frac{s^2}{2}}ds\,.$$

The last integral involves the so-called *error function* (integral of the Gaussian bell curve) that has no explicit expression in terms of elementary functions.

When the problem has no initial conditions we can find a whole family of solutions depending on an arbitrary constant. Going back to (3.7), we can integrate leaving an arbitrary constant on the right hand side:

$$(my)' = mq \Rightarrow my = \int mq\,dx + C$$

Solving for y:

$$y(x) = \frac{1}{m(x)} \left(\int mq \, dx + C \right) \tag{3.8}$$

where the function y defined in (3.8) is the *general solution* of the first order ordinary differential equation that has an arbitrary constant C. This constant will be fixed once we impose initial conditions. Notice that there is no need to take into account the integration constant in the primitive of p, because it cancels out when it is combined with the integral of the first term and can be melt with C as follows:

$$\frac{C}{m(x)} = Ce^{-(P(x)+c)} = Ce^{-c}e^{-P(x)} = \frac{Ce^{-c}}{m(x)} = \frac{\tilde{C}}{m(x)} .$$

So, the general solution depends on only one arbitrary constant.

4. Second order linear equations

4.1. Introduction

The general form of a second order differential equation is:

$$y'' = f\left(x, y, y'\right) . \tag{4.1}$$

Notice that Newton's law is precisely a second order differential equation:

$$my'' = F .$$

If we want to solve it we must know the functional form of the force F, that is, its dependence on time, position or velocity. Changing these functional forms we obtain different second order equations.

In general, the method for solving (4.1) depends of the form of the function f. A special case is the one in which the function f is *linear* on the variables y and y', that is, we can write the eqution as:

$$y'' + a_1\left(x\right) y' + a_0\left(x\right) y = h\left(x\right) .$$

In this particular case $f\left(x, y, y'\right) = -a_1\left(x\right) y' - a_0\left(x\right) y + h\left(x\right)$, that is, the function represents a plane for fixed x. When $h \equiv 0$ we say that the equation is *homogeneous*. An example of second order linear differential equation is:

$$y'' + e^x y' + \left(\sin x\right) y = 1$$

while

$$y'' + \left(y'\right)^2 = \cos x$$

is *non-linear*.

The general solution to a second order equation is a family of functions that depends on two free coefficients. These coefficients will be fixed once *initial or boundary conditions are prescribed*. Let us see an example:

$$y'' = 1 \Rightarrow y' = x + A \Rightarrow y = \frac{x^2}{2} + Ax + B .$$

The parameters A and B can be fixed with initial data:

$$y\left(0\right) = y_0 , \quad y'\left(0\right) = v_0$$

so we can solve for A:

$$v_0 = A\,,$$

and B

$$y_0 = B\,.$$

It is also possible to give boundary conditions. That means that we can provide the value of y in both extremes of the interval. For example: $y\left(0\right) = C_0$ and $y\left(1\right) = C_1$ (this is related to the so-called *shooting problem*). For this boundary value problem we have:

$$\begin{aligned} C_0 &= B\,, \\ C_1 &= \frac{\left(1\right)^2}{2} + A\,1 + B \Rightarrow A = C_1 - C_0 - \frac{1}{2}\,. \end{aligned}$$

In general, for a linear initial boundary value problem of order 2 we can provide the following existence theorem.

Theorem Given the second order differential equation

$$y'' + a_1\left(x\right)y' + a_0\left(x\right)y = h\left(x\right) \tag{4.2}$$

together with the initial conditions

$$y\left(x_0\right) = y_0 \qquad y'\left(x_0\right) = y_1\,,$$

if the functions a_0, a_1 and h *are continuous* in some interval (α, β) such that $x_0 \in (\alpha, \beta)$, then the equation has a unique solution y defined for all $x \in (\alpha, \beta)$.

Similarly to the first order linear equations, we can find the *general solution* of a linear second order differential equation finding a particular solution and the whole family of solutions to the homogeneous equation. That is, assume that by some method we were able to find a solution $y_p\left(x\right)$ of (4.2) and let $y\left(x\right)$ be the general solution of the equation (that is, with two free parameters). Is easy to check that the function $y_h = y - y_p$ *satisfies the associated homogeneous equation*:

$$\begin{aligned} y_h'' + a_1\left(x\right)y_h' + a_0\left(x\right)y_h &= \left(y - y_p\right)'' + a_1\left(x\right)\left(y - y_p\right)' + a_0\left(x\right)\left(y - y_p\right) \\ &= \left(y'' + a_1\left(x\right)y' + a_0\left(x\right)y\right) - \left(y_p'' + a_1\left(x\right)y_p' + a_0\left(x\right)y_p\right) \\ &= h\left(x\right) - h\left(x\right) \equiv 0\,. \end{aligned}$$

We conclude that if we find *one* particular solution y_p and we have the general solution y_h to the *homogeneous equation*, then

$$y = y_h + y_p$$

are *all* the solutions to the original equation.

We will focus in a very special set of second order differential equations, the so-called *constant coefficient equations* given by:

$$y'' + a_1 y' + a_0 y = h\left(x\right)$$

where a_1 and a_0 are arbitrary constants.

4.2. The homogeneous equation with constant coefficients

We will find the general solution of the following equation where a_1 y a_0 are constants:

$$y'' + a_1 y' + a_0 y = 0\,. \tag{4.3}$$

The analogous first order equation, that is the one given by:

$$y' + a_0 y = 0$$

has a solution in terms of an exponential function:

$$y\,(x) = Ke^{-a_1 x}\,.$$

The idea is to try a solution of (4.3) of the form:

$$e^{rx}\,,$$

with the idea of finding the values of r that turn e^{rx} into a solution. By inserting this expression in the equation, we find that r must be a solution of:

$$r^2 e^{rx} + a_1 r e^{rx} + a_0 e^{rx} = 0 \Rightarrow r^2 + a_1 r + a_0 = 0\,.$$

That is, r must be a root of the polynomial (called the *associated polynomial*) which are given by:

$$r = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0}}{2}\,.$$

Notice that these numbers may be complex. To extract solutions, we must consider two cases:

Real roots: $a_1^2 - 4a_0 \geq 0$

If the associated polynomial has two *different* roots ($a_1^2 - 4a_0 > 0$), denoted by r_1 and r_2, we will have two solutions of the differential equation given by:

$$e^{r_1 x} \qquad e^{r_2 x}\,.$$

In this case, the general solution of the differential equation is given by:

$$y_h\,(x) = c_1 e^{r_1 x} + c_2 e^{r_2 x}\,,$$

where we used that a linear combination of solutions is also a solution (you can prove this!). In this case we have two free parameters so we have all the possible solutions. If $r_1 = r_2$ ($a_1^2 - 4a_0 = 0$) we have a problem: we don't have two independen functions to combine, so we are not able to span the whole set of solutions of the equation. A natural way to find a second, independent solution is as follows: let us assume that

instead of having $r_1 = r_2$ we have that the roots are *very close each other*. That is, assume that

$$r_1 = r \quad r_\delta = r + \delta$$

where δ is a small number. In this case we have two solutions, as before:

$$e^{rx} \quad e^{r_\delta x}\,.$$

If we take a convenient linear combination we still have a solution to our equation:

$$y_\delta = \frac{1}{\delta}\left(e^{r_\delta x} - e^{rx}\right)\,,$$

then y_δ is a solution to the modified equation, but when $\delta \to 0$ we expect to find a solution of the original equation. Taking limit:

$$\lim_{\delta\to 0} y_\delta = \frac{1}{\delta}\left(e^{(r+\delta)x} - e^{rx}\right) = e^{rx}\lim_{\delta\to 0}\left(\frac{e^{\delta x}-1}{\delta}\right) = e^{rx}x\,.$$

We must now check that xe^{rx} is a solution to the original problem, that is, when $r_1 = r_2$. We calculate the derivatives:

$$\begin{aligned}(xe^{rx})'' + a_1\left(xe^{rx}\right)' + a_0 e^{rx} &= \left(e^{rx} + xre^{rx}\right)' + a_1\left(e^{rx} + xre^{rx}\right) + a_0 xe^{rx}\\ &= xe^{rx}\left(r^2 + a_1 r + a_0\right) + e^{rx}\left(a_1 + 2r\right)\,.\end{aligned}$$

Remember that, in this case we have that:

$$a_1^2 = 4a_0 \Rightarrow r = -\frac{a_1}{2}\,.$$

The term in the first parentheses is zero because r is a root of the polynomial. The second one is zero due to the previous expression. In that case we can say that the general solution to the homogeneous equation is given by:

$$y_h = c_1 e^{rx} + c_2 xe^{rx}$$

where c_1 and c_2 are two arbitrary constants.

Complex roots: $a_1^2 - 4a_0 < 0$

In this case we take advantage of the fact that the polynomial has *real coefficients.* So, both roots are conjugate of each other, that is:

$$r = -\frac{a_1}{2} \pm i\frac{\sqrt{a_1^2 - 4a_0}}{2} = \alpha \pm i\beta\,.$$

The solutions are given by

$$e^{(\alpha+i\beta)x} \quad \text{and} \quad e^{(\alpha-i\beta)x}\,.$$

We can find real functions by combining both solutions in the following way:

$$y_1(x) = \frac{e^{(\alpha+i\beta)x} + e^{(\alpha-i\beta)x}}{2} = e^{\alpha x}\cos(\beta x)$$

and

$$y_2(x) = \frac{e^{(\alpha+i\beta)x} - e^{(\alpha-i\beta)x}}{2i} = e^{\alpha x}\sin(\beta x)$$

Then, the general solution of the homogeneous equation is given in this case by:

$$y_h(x) = e^{\alpha x}(c_1\cos\beta x + c_2\sin\beta x)\ .$$

4.3. The non-homogeneous equation with constant coefficients

The method of indeterminate coefficients

We will solve a class of non-homogeneous equations using method that anticipates the possible solutions by looking at the form of the right hand side, that is, the function $h(x)$ in the following expression:

$$y'' + a_1 y' + a_2 y = h(x) \tag{4.4}$$

For example, if the function h is an exponential as

$$h(x) = e^{cx}\,,$$

we can try with a particular solution of the form:

$$y_p(x) = Ae^{cx}$$

and fix the value of the constant A in order to satisfy the equation. If we try this expression in (4.4) we have that A must satisfy:

$$y_p'' + a_1 y_p' + a_0 y = e^{cx} \Rightarrow c^2 Ae^{cx} + a_1 cAe^{cx} + a_0 Ae^{cx} = e^{cx}\,.$$

And then:

$$A\left(c^2 + a_1 c + a_0\right) = 1\,.$$

That is, if $c^2 + a_1 c + a_0 \neq 0$, we obtain a particular solution Ae^{cx} with:

$$A = \frac{1}{c^2 + a_1 c + a_0}\,.$$

If c *was a root* of the associated polynomial $(x^2 + a_1 x + a_0)$ we cannot clear the value of A and the former method is not valid. A similar case is found when we consider a first order differential equation such as:

$$y' - cy = e^{cx}\,.$$

If we try to find a particular solution following the same method, we will have the identity:

$$A\left(ce^{cx} - ce^{cx}\right) = e^{cx}\,,$$

so there is no value of A satisfying this relationship. In the case of first order (linear) differential equations we already studied the method of the integrating factor: multipliying both sides by the integrating factor $m(x) = e^{-cx}$ we obtain

$$e^{-cx}y' - ce^{-cx}y = 1 \Rightarrow \left(e^{-cx}y\right)' = 1\,.$$

Now we integrate directly:

$$e^{-cx}y = x + C \Rightarrow y = xe^{cx} + e^{cx}C\,.$$

If we take $C = 0$ we have a solution of the form:

$$y_p(x) = xe^{cx}\,.$$

That is, when c is a root of the associated polynomial, we can also try in (4.4) with a function of the form Axe^{cx} to obtain a particular solution. By replacing this expression in the equation we find that:

$$\begin{aligned}\left((Axe^{cx})'' + a_1\left(Axe^{cx}\right)' + a_0 Axe^{cx}\right) &= Ae^{cx}\left(\left(2c + c^2x\right) + a_1(1 + cx) + a_0x\right)\\ &= Ae^{cx}\left(x\left(c^2 + a_1c + a_0\right) + 2c + a_1\right)\,.\end{aligned}$$

The first parenthesis cancels out because c is a root of the polynomial. We have then the following identity:

$$Ae^{cx}(2c + a_1) = e^{cx} \Rightarrow A(2c + a_1) = 1\,.$$

If $2c + a_1 \neq 0$ the problem is solved and we can write:

$$A = \frac{1}{2c + a_1}\,.$$

But if $2c + a_1 = 0$ then c is a *double root* of the associated polynomial (that is, is a root of the polynomial and of its derivative), and we must try a different solution. We try then with a function of the form Ax^2e^{cx}. This function is found as follows (you don't need to understand this now, it is an explanation of why the particular solution is chosen in this way). If we perturb a bit the value of c by a small number δ, that is $c \to c + \delta$, we have that, after trying with this expression in the differential equation

$$y_\delta(x) = \frac{x}{\delta}\left(e^{(c+\delta)x} - e^{cx}\right)$$

we have a term of the form

$$\frac{x\left((c+\delta)^2 + a_1(c+\delta) + a_0\right) + 2(c+\delta) + a_1}{\delta}e^{(c+\delta)x}\,.$$

Taking into account that $2c+a_1 = 0$ and that c is a root of the associated polynomial, we have:

$$\frac{x\left(2c\delta + \delta^2 + a_1\delta\right) + 2\delta}{\delta} e^{(c+\delta)x} \underset{\delta\to 0}{\to} 2e^{cx}\,.$$

For the same limit in y_δ we obtain:

$$\lim_{\delta\to 0} y_\delta\left(x\right) = x^2 e^{cx}\,.$$

We can then try a particular solution of the form:

$$y_p\left(x\right) = Ax^2 e^{cx}\,.$$

After replacing this expression in the differential equation we have:

$$\begin{aligned}\left(Ax^2e^{cx}\right)'' + a_1\left(Ax^2e^{cx}\right)' + a_0Ax^2e^{cx} &= Ae^{cx}\left(\left(2 + 4cx + x^2c^2\right) + a_1\left(2x + x^2c\right) + a_0x^2\right)\\ &= Ae^{cx}\left(2 + \left(4c + 2a_1\right)x + \left(c^2 + a_1c + a_0\right)x^2\right)\\ &= 2Ae^{cx}\,,\end{aligned}$$

so, by choosing:

$$A = \frac{1}{2} \Rightarrow y_p\left(x\right) = \frac{x^2}{2}e^{cx}\,,$$

we have the particular solution that we were looking for.

In general, the method consists of looking for particular solutions that belong to the "same family" as $h\left(x\right)$. If they belong to the family of trigonometric functions, as sine and cosine, we can look for linear combinations of the form:

$$y_p\left(x\right) = A\cos x + B\,\mathrm{sen}x\,.$$

But, again, if this linear combination is a solution of the homogeneous equation, the method will not provide a useful solution. In that case we can try with:

$$y_p\left(x\right) = x\left(A\cos x + B\,\mathrm{sen}x\right)\,.$$

If we still are not able to find a particular solution, we can try again with a power two in x:

$$y_p\left(x\right) = x^2\left(A\cos x + B\,\mathrm{sen}x\right)\,.$$

The solution is almost the same when $h\left(x\right)$ is a *polynomial of degree n*. In that case we try with particular solutions of the form:

$$y_p\left(x\right) = A_0 + A_1x + \cdots + A_nx^n\,,$$

and we must find the coefficients A_i that satisfy the non-homogeneous equation.

4.4. Laplace transform

When the right hand side does not belong to a family of functions as exponentials, trigonometric or polynomials (or combinations of them), we can apply another method to find a particular solution. The method consist in applying an integral transformation that is called the *Laplace transform.* This turns the differential equation in an *algebraic* equation.

Definition Consider a function $y(t)$ for $t \geq 0$. The Laplace transform of y, is defined as:

$$\tilde{y}(s) \equiv \mathcal{L}[y](s) = \int_0^\infty y(t)\, e^{-st} dt\,,$$

where s is a real variable.

The most useful property, for solving differential equations, is given by the way that the derivatives are transformed. Consider the transform of $y'(t)$ combined with integration by parts:

$$\begin{aligned} \tilde{y'}(s) &= \int_0^\infty y'(t)\, e^{-st} dt = \lim_{A\to\infty} \left(y(A)\, e^{-sA} - y(0)\, e^{-s0} \right) + s \int_0^\infty y(t)\, e^{-st} dt \\ &= s\tilde{y}(s) - y(0)\ . \end{aligned}$$

Theorem The Laplace transform of the derivative of a function $y(t)$, $t \geq 0$ is given by:

$$\mathcal{L}[y'](s) = s\mathcal{L}[y](s) - y(0)\ .$$

This property is very important to find particular solutions to the equations with constant coefficients, because it turns differentiation into a *multiplication by s*.

Some examples of transformed functions:

$$\begin{aligned} \mathcal{L}[e^{bt}](s) &= \frac{1}{s-b} \qquad s > b\,, \\ \mathcal{L}[1](s) &= \frac{1}{s} \qquad s > 0\,, \\ \mathcal{L}[\cos \omega t](s) &= \frac{s}{s^2+\omega^2} \qquad s > 0\,, \\ \mathcal{L}[\sin\, \omega t](s) &= \frac{\omega}{s^2+\omega^2} \qquad s > 0\,, \\ \mathcal{L}[t^n](s) &= \frac{n!}{s^{n+1}} \qquad s > 0\,. \end{aligned}$$

We can say that the functions on the left (the ones to which we apply the transform) are the *inverse transforms* of the functions on the right hand side. For example:

$$\mathcal{L}^{-1}\left[\frac{1}{s}\right](t) = 1\,.$$

These *inverse transforms* are useful when we try to go back to the original variable and solving for the unkown function in a differential equation. Using a list of transforms, together with the following properties, we will be able to solve a broad set of equations:

L1 The Laplace transform is *linear,* in the following sense: given two functions $y_1(t)$, $y_2(t)$ and two constants c_1, c_2, we have that

$$\mathcal{L}[c_1 y_1 + c_2 y_2](s) = c_1 \mathcal{L}[y_1](s) + c_2 \mathcal{L}[y_2](s) \ .$$

L2 The transform of the derivative of a function is given by:

$$\mathcal{L}[y'](s) = s\mathcal{L}[y](s) - y(0) \ .$$

If we apply the property of the transform of a derivative several times, we reach the following identity:

$$\mathcal{L}[y^{(n)}](s) = s^n \mathcal{L}[y](s) - s^{n-1} y(0) - s^{n-2} y'(0) - \cdots - y^{(n-1)}(0)$$

where $y^{(n)}$ is the derivative of order n. In particular, for the second derivative we have that:

$$\mathcal{L}[y''](s) = s^2 \mathcal{L}[y](s) - sy(0) - y'(0) \ .$$

L3 The former property, (for the first derivative) can be inverted:

$$\mathcal{L}\left[\int_0^t y(\tau)\, d\tau\right](s) = \frac{1}{s}\mathcal{L}[y](s) \ .$$

L4 This property involves a *delay* in the transformed variable:

$$\mathcal{L}[e^{bt} y](s) = \mathcal{L}[y](s - b) \ .$$

L5 This is useful when computing the *inverse transform of derivatives:*

$$\mathcal{L}[t^n y](s) = (-1)^n \frac{d^n}{ds^n}\left(\mathcal{L}[y](s)\right) \ .$$

L6 The last is a *convolution property* (see the definition given in (2.15). Let $f = g * h(t)$, then:

$$\mathcal{L}[f](s) = \mathcal{L}[g](s) \cdot \mathcal{L}[h](s)$$

so we must have:

$$\mathcal{L}^{-1}[\tilde{g} \cdot \tilde{h}](t) = (g * h)(t) \ .$$

Remark: In this case we must extend $g(t)$ and $h(t)$ as *zero* for $t < 0$. Notice that the convolution is defined as an integral *in the whole* real line and Laplace transform is an integral over the half line.

Examples: a) Consider the equation

$$y'' - y = e^{-t} \cos(\omega t)$$

and compute a particular solution using Laplace transform.

Solution: (we write $\widetilde{y}\,(s)$ for $\mathcal{L}[y](s)$). Taking Laplace transform on both sides we obtain (we may assume $y\,(0)=0$ and $y'\,(0)=0$ because we are looking for *any* particular solution):

$$s^2\tilde{y}\,(s)-\tilde{y}\,(s)=\frac{s+1}{(s+1)^2+\omega^2}$$

where we applied the transform of the cosine and the property **L4**. We have then:

$$\left(s^2-1\right)\tilde{y}\,(s)=\frac{s+1}{(s+1)^2+\omega^2}\,.$$

Now we cancel a factor and write:

$$\tilde{y}\,(s)=\frac{1}{(s-1)\left((s+1)^2+\omega^2\right)}=\frac{A}{s-1}+\frac{B(s+1)+C}{(s+1)^2+\omega^2}$$

where we solve for the constants:

$$\begin{aligned}\frac{1}{(s-1)\left((s+1)^2+\omega^2\right)} &= \frac{A\left(s^2+2s+1+\omega^2\right)+B\left(s^2-1\right)+C\,(s-1)}{(s-1)\left((s+1)^2+\omega^2\right)}\\ &\Rightarrow A+B=0,\ 2A+C=0,\ A\left(1+\omega^2\right)-B-C=1\,.\end{aligned}$$

So

$$\begin{aligned}A\left(2+\omega^2\right)-C &= 1\\ 2A+C &= 0\end{aligned}$$

then

$$A=\frac{1}{\omega^2+4},\quad B=-\frac{1}{\omega^2+4}\quad C=-\frac{2}{\omega^2+4}\,.$$

So we have that:

$$\tilde{y}\,(s)=\frac{A}{s-1}+\frac{B(s+1)}{(s+1)^2+\omega^2}+\frac{C}{(s+1)^2+\omega^2}$$

Applying now inverse transform (using the table and the properties) we have that, a particular solution is given by:

$$y\,(t)=\frac{e^t}{\omega^2+4}-\frac{e^{-t}}{\omega^2+4}\left(\cos\left(\omega t\right)+\frac{2}{\omega}\sin\left(\omega t\right)\right)\,.$$

The first term is a solution of the homogeneous equation (so you can take it away and the function is still a particular solution).

b) Consider the following case involving initial conditions

$$\begin{cases} y''+5y'+6y=e^{-2t}\\ y(0)=1, y'(0)=0\end{cases}$$

we take Laplace transforms on both sides:

$$s^2\tilde{y}(s) - sy(0) - y'(0) + 5\left(s\tilde{y}(s) - y(0)\right) + 6\tilde{y}(s) = \frac{1}{s+2}.$$

Using the initial conditions we have that:

$$\left(s^2 + 5s + 6\right)\tilde{y}(s) = \frac{1}{s+2} + (s+5).$$

The roots of $s^2 + 5s + 6$ are given by: -2 and -3, so we can write:

$$\tilde{y}(s) = \frac{1}{(s+2)^2(s+3)} + \frac{s+5}{(s+2)(s+3)}$$

By using simple fractions it is possible to find constants A and B such that:

$$\frac{1}{(s+2)(s+3)} = \frac{A}{s+2} + \frac{B}{s+3}$$

so we have that $A = -B$ and $3A + 2B = 1$. So, $A = 1$, $B = -1$. We can now expand:

$$\begin{aligned}
\tilde{y}(s) &= \frac{1}{s+2}\left(\frac{A}{s+2} + \frac{B}{s+3}\right) + (s+5)\left(\frac{A}{s+2} + \frac{B}{s+3}\right) \\
&= \frac{A}{(s+2)^2} + \frac{B}{(s+2)(s+3)} + \frac{A(s+5)}{s+2} + \frac{B(s+5)}{s+3} \\
&= \frac{A}{(s+2)^2} + \frac{AB}{s+2} + \frac{B^2}{s+3} + A\left(1 + \frac{3}{s+2}\right) + B\left(1 + \frac{2}{s+3}\right)
\end{aligned}$$

Replacing the values of A and B we obtain:

$$\tilde{y}(s) = \frac{1}{(s+2)^2} + \frac{2}{s+2} - \frac{1}{s+3}.$$

All the expressions have inverse transforms that can be found in the table, so:

$$y(t) = te^{-2t} + 2e^{-2t} - e^{-3t}.$$

Bibliography

[1] *Applied Mathematics: Body & Soul.* Kenneth Eriksson, Don Estep and Claes Johnson, Vol 1. Freely available in the web: http://www.csc.kth.se/~cgjoh/eriksson-vol1.pdf

[2] *Introduction to Real Analysis,* Robert G. Bartle and Donald Sherbert. John Wiley & Sons, Inc. 2011 (4th edition).

[3] *Partial Differential Equations of Mathematical Physics and Integral Equations*, Ronald Guenther and John Lee. Prentice Hall 1988.

[4] *A first course in differential equations*, Frank G. Hagin. Prentice Hall 1975.

Part II.

Algebra

5. Integers

Bibliography: Al*gebra for Computer Science*, Lars Gårding, Torbjörn Tambour. Universitext, Springer 1988. *Abstract Algebra: Theory and Applications,* Thomas Judson. Free download: http://abstract.ups.edu/download/aata-20150812-print.pdf.

5.1. The induction principle

Natural numbers are symbols used for counting:

$$\mathbb{N} = \{1, 2, \dots\} \ .$$

We may include the number zero in this set, but it is more *natural* to start by 1 if we want to count things. If we include 0 we use the notation $\mathbb{N}_0$.

The set of integer numbers is denoted by $\mathbb{Z}$ including zero and the negative numbers:

$$\mathbb{Z} := \{\dots, -2, -1, 0, 1 \dots\} \ .$$

We will refer to $\mathbb{N}$ as the set of *positive integers.*

First Principle of Mathematical Induction. Let $S(n)$ be a statement for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some $n_0 \in \mathbb{N}$. If for all integers k with $k \geq n_0$, $S(k)$ implies that $S(k+1)$ is true, then $S(n)$ is true for all integers n greater than or equal to n_0.

An equivalent formulation is as follows:

Second Principle of Mathematical Induction. Let $S(n)$ be a statement for $n \in \mathbb{N}$ and suppose that $S(n_0)$ is true for some $n_0 \in \mathbb{N}$. If $S(n_0)$, $S(n_0+1), \dots, S(k)$ imply that $S(k+1)$ is true for $k \geq n_0$, then the statement $S(n)$ is true for all numbers $n \geq n_0$.

Definition A nonempty subset S of $\mathbb{Z}$ is *well-ordered* if S contains a *least element.* Notice that the set $\mathbb{Z}$ is not well-ordered since it does not contain a smallest element. However, the natural numbers are well-ordered.

Principle of Well-Ordering. Every nonempty subset of the natural numbers is well-ordered.

The Principle of Well-Ordering is equivalent to the Principle of Mathematical Induction:

Theorem The Principle of Mathematical Induction implies the Principle of Well-Ordering. That is, every nonempty subset of $\mathbb{N}$ contains a least element. Reciprocally, the Principle of Well-Ordering implies the Principle of Mathematical Induction.

Proof (adapted from Judson, thm. 2.2) If a set $A \subset \mathbb{N}$ contains the number 1, it is obviously well ordered, since $1 \leq a$ for all $a \in A$. Now let $S(n)$ be the statement: "if a subset of natural numbers contains n, it is well ordered". Assume that $S(k)$ is true for $1 \leq k \leq n$, and we will show that then $S(n+1)$ must be true. Take a set A containing the number $n+1$. If $n+1$ is the least element, the statement is proved. If $n+1$ is not the least element, there must be some $1 \leq k \leq n$ such that $k \in A$. Then, A is well ordered by inductive assumption and $S(n)$ must be true for all n. Every nonempty subset of $\mathbb{N}$ contains at least one number n, then it must be well ordered because $S(n)$ is true for all n.

Now assume that the PWO is valid and we will show that PMI is true. Assume that statement $S(1)$ is true, and that $S(n) \Rightarrow S(n+1)$ for all n. Consider the subset A of natural numbers containing the values of n such that $S(n)$ is not true. If A is empty, then the PMI is true. If A is not empty, then by the PWO there must be a least value $a_0 \in A$, and a_0 must be greater than 1. So, for $n = a_0 - 1$, $S(n)$ must be true. Then, $S(n+1) \equiv S(a_0)$ must also be true. Then, A must be empty, and $S(n)$ is true for all $n \in \mathbb{N}$.■

Well ordering also implies the division algorithm:

Theorem (Division Algorithm) Let a and b be integers, with $b > 0$. Then there exist unique integers q and r such that

$$a = bq + r$$

where $0 \leq r < b$.

Proof We must first prove that the numbers q and r actually exist. Then we must show that if q' and r' are two other such numbers, then $q = q'$ and $r = r'$.

Existence of q and r. Let

$$S = \{a - bk : k \in \mathbb{Z} \text{ and } a - bk \geq 0\}.$$

If $0 \in S$, then b divides a, and we can let $q = a/b$ and $r = 0$. If $0 \notin S$, we can use the Well-Ordering Principle. We must first show that S is nonempty. If $a > 0$, then $a - b \cdot 0 \in S$. If $a < 0$, then $a - b(2a) = a(1 - 2b) \in S$. In either case $S \neq \emptyset$. By the Well-Ordering Principle, S must have a smallest member, say $r = a - bq$. Therefore, $a = bq + r$, $r \geq 0$. We now show that $r < b$. Suppose that $r > b$. Then

$$a - b(q+1) = a - bq - b = r - b > 0.$$

In this case we would have $a - b(q+1)$ in the set S. But then $a - b(q+1) < a - bq$, which would contradict the fact that $r = a - bq$ is the smallest member of S. So $r \leq b$. Since $0 \notin S$, $r \neq b$ and so $r < b$.

Uniqueness of q and r. Suppose there exist integers r, r', q, and q' such that

$$a = bq + r, 0 \leq r < b \qquad \text{and} \qquad a = bq' + r', 0 \leq r' < b.$$

Then $bq + r = bq' + r'$. Assume that $r' \geq r$. From the last equation we have $b(q - q') = r' - r$; therefore, b must divide $r' - r$ and $0 \leq r' - r \leq r' < b$. This is possible only if $r' - r = 0$. Hence, $r = r'$ and $q = q'$. ■

Greatest common divisor

Definitions An integer d is called a *common divisor* of a and b if $d \mid a$ and $d \mid b$. The *greatest common divisor* of integers a and b is a positive integer d such that d is a common divisor of a and b and if d' is any other common divisor of a and b, then $d' \mid d$. We write $d = \gcd(a, b)$. We say that two integers a and b are *relatively prime* or *coprime* if $\gcd(a, b) = 1$. A positive integer m is called the *least common multiple* of a and b if it is the least natural m such that $a|m$ y $b|m$. We write $m = \operatorname{lcm}(a, b)$.

Theorem BÉZOUT IDENTITY Let a and b be nonzero integers. Then there exist integers x and y such that

$$\gcd(a, b) = ax + by.$$

Furthermore, the greatest common divisor of a and b is unique.

Proof Let

$$S = \{am + bn : m, n \in \mathbb{Z} \text{ and } am + bn > 0\}.$$

The set S is nonempty, just take $m = \text{sig}a$ and $n = \text{sig}b$ and $am + bn$ belongs to S; hence, by the Well-Ordering Principle S must have a smallest member, say $d = ax + by$. We claim that $d = \gcd(a, b)$. Write $a = dq + r'$ where $0 \leq r' < d$. If $r' > 0$, then

$$\begin{aligned} r' &= a - dq \\ &= a - (ax + by)q \\ &= a - axq - byq \\ &= a(1 - xq) + b(-yq), \end{aligned}$$

which is in S. But this would contradict the fact that d is the smallest member of S. Hence, $r' = 0$ and d divides a. A similar argument shows that d divides b. Therefore, d is a common divisor of a and b.
Suppose that d' is another common divisor of a and b, and we want to show that $d' \mid d$. If we let $a = d'h$ and $b = d'k$, then

$$d = ax + by = d'hx + d'ky = d'(hx + ky).$$

So d' must divide d. Hence, d must be the unique greatest common divisor of a and b. ■

Corollary Let a and b be two integers that are relatively prime. Then there exist integers x and y such that $ax + by = 1$.

Euclidean algorithm

Let a, b be integer numbers with $a > b > 0$. Euclid's algorithm is used to find the greatest common divisor of a and b, and proceeds as follows:

$$\gcd(a, b) \to \gcd(b, r)$$

where r is the remainder of the integer division of a by b, i.e. $a = qb + r$. If $c := \gcd(a, b)$, is easy to see that c must divide r, because as it must divide a and b:

$$\underbrace{q'c}_{a} = q\,\underbrace{q''c}_{b} + r \Rightarrow c\,(q' - qq'') = r\,.$$

The number between parentheses is non-negative, because $a - qb = r \geq 0$. Then $\gcd(b, r) \geq c$. The remainder r is strictly less than b, so the algorithm has an end in a finite number of steps. On the other side, as any number that divides r and b divides a too, then $\gcd(b, r) \leq \gcd(a, b) = c$, then $\gcd(b, r) = c$. The last nonzero remainder is the $\gcd(a, b)$. Taking the inverse steps of the algorithm we find integer numbers x and y such that

$$xa + yb = c\,.$$

Steps of the algorithm

- We start with the pair (a, b). We compute the integer division: $a = bq_1 + r_1$. Notice that r_1 is a linear combination of a and b: $r_1 = a - q_1 b$. We can write this as $r_1 = x_1 a + y_1 b$, where $x_1 = 1$, $y_1 = -q_1$.
- We continue with the pair (b, r_1). We compute: $b = r_1 q_2 + r_2$. Then, r_2 is a linear combination of b and r_1: $r_2 = b - r_1 q_2$. So $r_2 = b - q_2\,(x_1 a + y_1 b)$, then $r_2 = (-q_2 x_1)\,a + (1 - q_2 y_1)\,b$, that is $x_2 = -q_2 x_1$ and $y_2 = 1 - q_2 y_1$.
- We continue computing $\gcd(r_{n-1}, r_n)$. r_{n+1} is a linear combination of r_{n-1} and r_n. That is: $r_{n+1} = r_{n-1} - q_{n+1} r_n$. So $r_{n+1} = (x_{n-1} a + y_{n-1} b) - q_{n+1}\,(x_n a + y_n b)$, so $x_{n+1} = x_{n-1} - q_{n+1} x_n$ and $y_{n+1} = y_{n-1} - q_{n+1} y_n$.
- When we reach $(r_N, 0)$, we have that r_N is the greatest common divisor of a and b. As r_N is a linear combination of the former remainders (r_{N-1} y r_{N-2}) and they are linear combination of a and b, ie, $r_{N-1} = x_{N-1} a + y_{N-1} b$ and $r_{N-2} = x_{N-2} a + y_{N-2} b$, we can write r_N as a linear combination of a and b:

$$r_N = x_N a + y_N b\,.$$

Example (Exercise 1 from Integers) Compute the greatest common divisor between $a = 10672$, and $b = 4147$.
We apply the algorithm, first step:

$$10672 = 4147 \times 2 + 2378$$

then $q_1 = 2$, $r_1 = 2378$. The following steps.
1. $a - 2b = 2378 = r_1$.
2. $b - r_1 = 1769 = r_2$.
3. $r_1 - r_2 = 609 = r_3$.
4. $r_2 - 2r_3 = 551 = r_4$.

5. $r_3 - r_4 = 58 = r_5$.
6. $r_4 - 9 \times r_5 = 29 = r_6$.
7. $r_5 - 2 \times r_6 = 0$,

so that, the gcd is $r_6 = 29$. To compute a Bézout identity, we can go backwards from step 6:

$$\begin{aligned} 29 &= r_4 - 9\,(r_3 - r_4) = 10r_4 - 9r_3 = 10\,(r_2 - 2r_3) - 9r_3 = -29r_3 + 10r_2 = -29\,(r_1 - r_2) + 10r \\ &= -29r_1 + 39r_2 = -29r_1 + 39\,(b - r_1) = -68r_1 + 39b = -68\,(a - 2b) + 39b = -68a + 175b. \end{aligned}$$

or we can also keep the values of x_i, y_i as follows:

1. $r_1 = a - 2b = 2378 \Rightarrow x_1 = 1, y_1 = -2$.
2. $r_2 = b - r_1 = 1769 \Rightarrow r_2 = b - (x_1 a + y_1)\,b \Rightarrow x_2 = -1, y_2 = 3$.
3. $r_3 = r_1 - r_2 = 609 = a\,(x_1 - x_2) + b\,(y_1 - y_2) \Rightarrow x_3 = 2, y_3 = -5$.
4. $r_4 = r_2 - 2r_3 = 551 = -5a + 13b \Rightarrow x_4 = -5, y_4 = 13$.
5. $r_5 = r_3 - r_4 = 58 = 7a - 18b \Rightarrow x_5 = 7, y_5 = -18$.
6. $r_6 = r_4 - 9 \times r_5 = 29 \Rightarrow -68a + 175b$.
7. $r_5 - 2 \times r_6 = 0$, obtaining the same result.

5.2. Primes and factorization

Definition Given a, b and c integers such that $a = bc$ we say that both b and c *divide* a, and are *divisors* of a. On the other hand, a is a multiple of b and of c. Sometimes, when a number x divides y we write it as: $x|y$.

Every integer a such that $a \neq \pm 1$ has at least four trivial divisors: ± 1 and $\pm$itself. These are called *trivial divisors*. If $a = \pm 1$ then it has two trivial divisors.

A positive integer is a *prime* if admits exactly two positive trivial divisors. The first primes are:

$$2, 3, 5, 7, 11, 13, 17, \ldots$$

Notice that 1 *is not prime* because we asked for *exactly two positive trivial divisors,* and 1 has only one.

Theorem Every integer $N > 1$ can be factorized as ap, where a is a positive integer and p is prime.

Proof We apply induction over N. For $N = 2$ we have that $2 = 1 \times 2$. Assume that the statement is true for $n \leq N - 1$ and we will prove it for $n = N$. The inductive hypothesis tells us that every number n such that $1 < n < N$ can be written as ap. If N does not have non trivial divisors, then N itself is prime and we conclude choosing $a = 1$ and $p = N$. If N is not a prime, it must admit nontrivial divisors, i.e.

$$N = ab\,.$$

As a and b are different from 1 and N, they must be less than N and greater than 1. By inductive hypothesis we have that $a = cp$ where c is natural and p is prime. Thus $N = \tilde{a}p$, where $\tilde{a} = bc$.■

Corollary Every positive integer $N > 1$ admits a factorization

$$N = p_1 \dots p_k \quad k \geq 1\,,$$

where p_i are primes.

Proof Again by induction, the case $N = 2$ is trivial. Assume that the factorization is valid for every $2 \leq n < N$. By the theorem, N admits a factorization $N = ap$ with p prime and a natural. If $a = 1$ we have that $N = p$ and the proof is finished. If $a > 1$ then it is a natural number less than N, that admits a factorization: $a = p_1 \dots p_k$, multiplying by p we obtain the factorization for N:

$$N = p_1 \dots p_k p\,.$$

■

The following theorem can be found in Euclid's Elements (300 BC).

Theorem There are infinite prime numbers.

Proof We will show that, for any list of prime numbers, there is always one more that is not in the list. Let p_1, p_2,..., p_n be primes. Consider the number

$$p_1 p_2 \dots p_n + 1.$$

This is an integer greater than 1, so we can factor it as

$$ap = p_1 p_2 \dots p_n + 1\,,$$

with a being a positive integer and p prime. If $a = 1$ then $p_1 p_2 \dots p_n + 1$ is prime and different from each p_k for all $k = 1, \dots, n$. If $a \neq 1$ and $p = p_k$ for some k, we would have that

$$ap - p_1 p_2 \dots p_n = 1\,,$$

then p would be a divisor of 1 (we may take it as a common factor). The only positive divisor of 1 is 1 itself, so p would not be prime. We conclude that p cannot be in the former list. ■

Property If p is prime and divides ab, then p divides a, or p divides b.

Proof Let us assume that p does not divide a. In this case, $\gcd(a, p) = 1$, because the only divisors of p are the trivial ones. Using Bézout's identity, there exists integers x, y such that:

$$xa + yp = 1\,.$$

By multiplying both sides by b we obtain:

$$xab + ypb = b\,,$$

using that p divides ab we have that: $ab = qp$. Putting together the last two identities we have that:

$$xqp + ypb = b \Rightarrow p\,(xq + yb) = b\,,$$

so that p must divide b. Concluding, if p does not divide a, then it must divide b (it may divide both numbers or only one of them).■

Definition The numbers defined by p^n for p prime and n a natural number are called *primary numbers.*

Theorem FUNDAMENTAL THEOREM OF ARITMETIC Every integer $N > 1$ can be factorized as a product of prime numbers. Such factorization is unique up to rearangement of the factors.

Proof We already proved that an integer greater than 1 can be factorized with prime numbers. So we will prove uniqueness. To show uniqueness we will use induction on N. The theorem is certainly true for $N = 2$ since in this case N is prime. Now assume that the result holds for all integers n such that $1 \leq n < N$, and

$$N = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l,$$

where $p_1 \leq p_2 \leq \cdots \leq p_k$ and $q_1 \leq q_2 \leq \cdots \leq q_l$. By a previous Corollary, $p_1 \mid q_i$ for some $i = 1, \ldots, l$ and $q_1 \mid p_j$ for some $j = 1, \ldots, k$. Since all of the p_i's and q_i's are prime, $p_1 = q_i$ and $q_1 = p_j$. Hence, $p_1 = q_1$ since $p_1 \leq p_j = q_1 \leq q_i = p_1$. By the induction hypothesis,

$$n' = p_2 \cdots p_k = q_2 \cdots q_l$$

has a unique factorization. Hence, $k = l$ and $q_i = p_i$ for $i = 1, \ldots, k$. ■

Property The Least Common Multiple (LCM) of two natural numbers a and b is given by:

$$\operatorname{lcm}(a, b) = \frac{a \cdot b}{\gcd(a, b)}$$

where $\gcd(a, b)$ is the greatest common divisor of a and b.

Proof The greatest common divisor contains all the common factors of a and b, so that $a = \gcd(a, b) \cdot x$ and $b = \gcd(a, b) \cdot y$, where x and y do not have common prime factors (they are coprime). Then $a \cdot b = \gcd(a, b)^2 \cdot x \cdot y$. Dividing both terms by $\gcd(a, b)$, we are cleraring the common factors so they appear in the minimum amount:

$$\frac{a \cdot b}{\gcd(a, b)} = \gcd(a, b) \cdot x \cdot y$$

So, the right hand side has all the factors of a and also of b, that is:

$$\gcd(a, b) \cdot x \cdot y = a \cdot y = b \cdot x\,.$$

As x and y are coprimes, this is the least number with the desired property. ■

5.3. Linear diophantine equations

Definition A *diophantine equation* is an equation in two or more unknowns such that only the integer solutions are searched or studied. The *linear diophantine equation* in two unknwons is given by:

$$ax + by = c \tag{5.1}$$

with $a, b, c \in \mathbb{Z}$, where x and y are the unknowns.

Proposition Equation (5.1) has integer solutions if and only if $d := \gcd(a, b)$ divides c.

Proof: If x_0, y_0 is an integer solution of (5.1), then clearly d divides c:

$$ax_0 + by_0 = a'dx_0 + b'dy_0 = d\,(a'x_0 + b'y_0) = c\,.$$

On the other side, if d divides c, we can write (dividing both sides of (5.1) by d):

$$a'x + b'y = c'\,,$$

where now $\gcd(a', b') = 1$. Using Bézout identity, there exists an integer solution u, v to the equation:

$$a'u + b'v = 1\,.$$

Finding u, v by means of Euclid's algorithm, and then multiplying by c' we have a solution of the equation:

$$a'(uc') + b'(vc') = c'\,.$$

■

The whole set of solutions Given two solutions (x, y) , (x_0, y_0) of a linear diophantine equation as in (5.1), we have that:

$$\begin{aligned} xa + yb &= c \\ x_0a + y_0b &= c \end{aligned}$$

then, taking the difference:

$$(x - x_0)a + (y - y_0)b = 0\,.$$

And then dividing by $d = \gcd(a, b)$ we have that

$$(x - x_0)\,a' + (y - y_0)b' = 0$$

As $\gcd(a', b') = 1$, we have that a' divides $(y - y_0)$ and b' divides $(x - x_0)$, so:

$$\gamma = \frac{x - x_0}{a'} = -\frac{y - y_0}{b'} \Rightarrow x = x_0 + \gamma a', y = y_0 - \gamma b'\,,$$

and these are all the possible solutions to the linear diophantine equation, by giving values to the integer γ. The fact that a' divides $y - y_0$ means that all the numbers y can be obtained from y_0 and a multiple of a':

$$y = y_0 + ka'.$$

This means that y is *congruent* to y_0 modulo a'.

5.4. Congruences

Definition Let m be a fixed integer. We say that x is *congruent* with y *modulo* m if m divides $x - y$. We denote this as follows:

$$x \equiv y \bmod m \text{ or } x \equiv_m y\,.$$

Every integer congruent with x modulo m form an *equivalence class* $x + \mathbb{Z}m$ denoted by $[x]_m \in \mathbb{Z}_m$.

The operations "modulo m" take as elements the equivalence classes. An integer number that belongs to a congruence class, *represents* that class.

Property There are exactly $|m|$ congruence classes modulo m.

Proof: Given an integer x, by the division algorithm there exist unique integers q and r such that $x = q|m| + r$, with $0 \leq r < |m|$. Each value of r defines a unique class, so there are only $|m|$ possible classes. ■

Property If $x \equiv_m y$ and $u \equiv_m v$ then $x \pm u \equiv_m y \pm v$ and $xu \equiv_m yv$.

Proof $x = y + qm$ and $u = v + km$ so we have that

$$x \pm u = (y \pm v) \pm (q + k)\, m,$$

that is, $x \pm u$ is in the same class as $y \pm v$. For the product:

$$\begin{aligned} xu &= (y + qm)\,(v + km) = yv + (vq + yk)\, m + qkm^2 \\ &= yv + (vq + yk + qkm)\, m \end{aligned}$$

so that xu and yv belong to the same class■

If $[x]_m$ is the congruence class of x modulo m, the former property shows that:

$$[x]_m + [y]_m = [x + y]_m \quad [x]_m\,[y]_m = [xy]_m\,.$$

This shows that $[0]_m$ is the *additive neutral element* and $[1]_m$ is the *multiplicative neutral element.* The two operations have further properties that turn $\mathbb{Z}_m$ into a *ring.*

Property Given an integer a, there exists an integer b such that

$$ab \equiv_m 1$$

if and only if $\gcd(a, m) = 1$, that is, if they are coprimes.

Proof If b is a solution of the congruence equation, we must have a number q such that

$$ab - 1 = qm \Rightarrow ab - qm = 1$$

then a is coprime with m, because if $d|a$ and $d|m$ it must divide 1, so $d = 1$. On the other side, if they are coprime, Bézout's identity provides numbers x and y such that

$$xa + ym = 1\,,$$

so that we can take $b = x$ ■

The number b with the previous property is called the *inverse* of a modulo m. The inverse modulo m is unique (as a class), because if there is another c with the same property, we must have:

$$a\,(b-c) \equiv_m 0\,.$$

As a and m are coprime, we have that $b-c \equiv_m 0$, that is, they are in the same class. If a amd m are coprime, the congruence

$$ax \equiv_m b$$

has a unique solution given by (the inverse is taken in the classs modulo m):

$$[x]_m = [a]_m^{-1}[b]_m\,.$$

Solving diophantine equations using congruences

There is a practical method to solve diophantine equations, where we can take advantage of operations between congruent numbers. Assume that, after transforming a given diophantine equation, que want to solve:

$$xm + yn = c\,,$$

where m and n are coprime. Instead of finding first a Bézout identity, we proceed as follows. First, taking into account that all the solutions for x are congruent modulo n we consider the given diophantine equation, modulo n:

$$xm + yn = c \pmod{n}$$

using now that

$$yn = 0 \pmod{n}$$

we find that

$$xm = c \pmod{n}$$

Now we solve using the inverse of m modulo n:

$$x = m^{-1}c \pmod{n}$$

where we understand that m^{-1} is *a representative* of the class modulo n that is the inverse of m. Let us call $x_0 = m^{-1}c$, so the complete set of solutions for x is given by:

$$x = x_0 + kn$$

Then, we can return to the diophantinte equation to clear for y:

$$(x_0 + kn)\,m + yn = c \Rightarrow y = \frac{c - (x_0 + kn)\,m}{n} = \frac{c - x_0 m}{n} - km$$

The first term is an integer, due to the fact that $x_0 m$ is congruent with c modulo n.

Simultaneous congruences

Consider the following congruences, for an unkown integer x:

$$x \equiv_{m_1} a_1 \quad x \equiv_{m_2} a_2 \quad \dots x \equiv_{m_n} a_n \,,$$

where the numbers m_i are pairwise coprime, that is $(m_i, m_j) = 1$ for $i \neq j$. We can solve these congruences one by one. Take the first:

$$x = a_1 + y m_1$$

and replace this relationship in the following equations, for $1 < i \leq n$ we have that:

$$a_1 + y m_1 \equiv_{m_i} a_i \Rightarrow y m_1 \equiv_{m_i} a_i - a_1 \,.$$

As m_1 is coprime with m_i for every i, we can take the inverse of m_1 modulo m_i:

$$y \equiv_{m_i} m_1^{-1} (a_i - a_1) \,,$$

obtaining $n-1$ congruences for each y. Thus, applying the method n times and going back we may compute x. The result will be given modulo $m_1 \times m_2 \cdots \times m_n$.

We can solve all the congruences in one step, using the following result.

Theorem CHINESE REMAINDER THEOREM. Given $\{m_k\}_{k=1,\dots n}$ coprime by pairs, $M = m_1 \dots m_n$ and $M_k = M/m_k$. Clearly, M_k and m_k are coprime for every $1 \leq k \leq n$. Consider the numbers c_k that are solutions of the following n congruences:

$$M_k c_k \equiv_{m_k} a_k \,.$$

Then, the number x given by:

$$x := M_1 c_1 + \cdots + M_n c_n \tag{5.2}$$

is the solution of the system of congruences.

Proof We will show that the so-defined number solves all the congruences. Given i, we must check that x satisfies:

$$M_1 c_1 + \cdots + M_n c_n \equiv_{m_i} a_i \quad \forall i \,.$$

Consider the diference:

$$\begin{aligned} x - a_i &= \sum_{k \neq i} M_k c_k + (M_i c_i - a_i) \\ &= m_i \sum_{k \neq i} M'_k c_k + q_i m_i \\ &= m_i \left(\sum_{k \neq i} M'_k c_k + q_i \right) . \end{aligned}$$

The first term M_k with $k \neq i$ has a factor m_i, then we can take m_i as a common factor, and $M'_k = M/(m_k m_i)$ (this is the product of all the m's, without m_k and m_i). In the last term we use the fact that $M_i c_i$ is congruent with a_i. ■

A sligthly more difficult problem can be solved too when we replace the equation $x \equiv_{m_i} a_i$ by $b_i x \equiv_{m_i} a_i$ where b_i, m_i are coprime. In this case the c_i solve:

$$b_i M_i c_i \equiv_{m_i} a_i$$

because $b_i M_i$ is still coprime with m_i. The solution has the same representation as in (5.2).

Uniqueness: The solution is *unique modulo* M, because if y is another solution, we have that $x - y \equiv_{m_i} 0$ for every i. Then the difference is divisible by M, so that

$$x \equiv_M y\,.$$

Example (Exercise 11 b) Solve:

$$x \equiv_3 2\,,\ x \equiv_4 3\,,\ x \equiv_5 4.$$

In this case $M = 3 \times 4 \times 5 = 60$ and the M_i are, respectively, 20, 15, 12. We have to solve then:

$$20c_1 \equiv_3 2 \quad 15c_2 \equiv_4 3 \quad 12c_3 \equiv_5 4\,.$$

The solution is $c_1 = 1$, $c_2 = 1$, $c_3 = 2$, so that

$$x = 20 + 15 + 24 = 59\,,$$

and is unique up to multiples of 60.

Example ("Desafío Matemático El País" 17/12/2013) We want to buy a lotery number (5 digits) that satisfy the following conditions:
1. All its digits are different.
2. If we enumerate the months of the year from 1 to 12, in any month of the year, after substracting to our number the number corresponding to the previous month, the result is divisible by the number of the month in which we presently are. This happens for all the months of the year.
SOLUTION: If x is the unkown number, and m are the months of the year ($m = 1, 2, \ldots, 12$) x has the property

$$x - (m-1) = q_m m$$

or what is the same:

$$x - (m-1) \equiv_m 0 \Rightarrow x + 1 - m \equiv_m 0$$

but m *is* congruent with 0 modulo m so we look for x such that

$$x + 1 \equiv_m 0\,, \quad m = 1, \ldots, 12.$$

So, m divides $x + 1$ for each m so, if M is the least common multiple of the numbers from 1 to 12, we must have that:

$$x + 1 \equiv_M 0\,.$$

The least common multiple M is given by: $2^3 \times 3^2 \times 5 \times 7 \times 11 = 27720$. Then, x is a number of the form:

$$x = q \times 27720 - 1\,.$$

If $q = 1, 2$ the digits are repeated. If $q = 3$ we have 83159, that satisfies all the conditions.

When b and m are not coprime

Up to now we solved the case in which $bx \equiv_m a$ with $(m, b) = 1$. What happens if they are not coprime? The congruence means:

$$bx - a = qm \Rightarrow xb - qm = a\,.$$

Notice that if d divides b and m, then it must divide a. In fact, if d greatest common divisor between b and m, it must divide a too. *This is a necessary and sufficient condition for the existence of a solution.* We can transform the problem in:

$$xb' - qm' = a' \qquad (b'x \equiv_{m'} a')\ .$$

where we divided b, m and a by the common factor d. Now we have that, b' and m' are coprime, and the system has a solution by finding the inverse of b' and then multiplying by a', that is:

$$x = (b')^{-1}\, a' \pmod{m'}\ . \tag{5.3}$$

Here we obtain *a class of solutions* of the original problem. As a class, is a unique solution of the "coprimized" problem, and now we may obtain all the solutions to the original problem (módulo m). Given x_0 a solution of (5.3) we have that:

$$x = x_0 + km'$$

are solutions of the initial problem too. Even though this numbers are all in the class m', some of them are *different* if we consider them as numbers in the congruence class m. Then, all the numbers satisfying

$$0 \leq x_0 + km' < m$$

for integer k, are different solutions to the initial problem. Notice that if m' is m/d where $d = \gcd(b, m)$ then

$$0 \leq x_0 + km' < dm'\,.$$

Then, if we take x_0 as the representative satisfying $0 \leq x_0 < m'$, there are different solutions (modulo m) for $k = 0, 1, 2, \ldots, (d-1)$. That is, there are exactly d solutions to the initial equation.

When the m_i's are not coprime

The Chinese Remainder Theorem gives solutions to the problem of simultaneous congruences

$$x \equiv_{m_1} a_1 \, , x \equiv_{m_2} a_2 \, , \ldots x \equiv_{m_n} a_n \, .$$

in the cases where m_i are pairwise coprime. When the m_i have common factors, we must change to an equivalent problem, after checking compatibility. Let us see two transformations that can be applied independently to each congruence.

Assume that we want to solve:

$$x \equiv_m a \, , \quad m = m_1 m_2 \text{ with } (m_1, m_2) = 1 \, .$$

In this case we have that

$$x - a = q m_1 m_2 \, ,$$

that is, $x - a$ is divisible by m_1 and also by m_2, so that we must have a couple of congruences:

$$x \equiv_{m_1} a \, , \quad x \equiv_{m_2} a \, .$$

On the other hand, is these congruences are simultaneously satisfied, then $x - a$ is divisible by m_1 and m_2 too, and as they are coprime, the product $m_1 m_2$ *must divide* $x - a$ (this is not valid if they have common factors, for example 2 divides 10 and 10 divides 10, but 2×10 does not divide 10).

So we showed that the initial congruence is equivalent to two congruences with coprime modules:

$$\textbf{rule1}\text{: if } (m_1, m_2) = 1 \text{ and } m = m_1 m_2 \Rightarrow x \equiv_m a \Leftrightarrow \begin{cases} x \equiv_{m_1} a \, , \\ x \equiv_{m_2} a \, . \end{cases}$$

Then, if we must solve two congruences such that m_1 has a common factor m with another m_2, we can replace each congruence by two.

After applying rule 1 we will find a problem of the form:

$$x \equiv_m a_1 \qquad x \equiv_m a_2$$

(in the former problem we had the same a and different m's, in this the a's are changed and the m is mantained). This implies that

$$x - a_1 = qm \qquad x - a_2 = km$$

that is, x is in the same class as a_1 and a_2 modulo m. Then a_1 *must be in the same class that* a_2 (mod m). This is easy to see by substracting the equations

$$a_2 - a_1 = (q - k)m \Rightarrow a_2 \equiv_m a_1 \, .$$

If this is not valid, *it is not possible to find a solution* of the simultaneous congruences. If the condition is true, then we can remove one of the two congruences, because they are redundant. We can state the rule as follows:

$$\textbf{rule2}\colon\ x \equiv_m a_1\,, x \equiv_m a_2 \Leftrightarrow \begin{cases} \text{if } a_1 \not\equiv_m a_2 & \text{no solution}\,, \\ \text{if } a_1 \equiv_m a_2 & x \equiv_m a_1\,. \end{cases}$$

With these two rules we can tranform a problem of simultaneous congruences by another that can be solved (if it is compatible) by the Chinese Remainder Theorem (CRT).

Example Exercise 11 c) has the following set of congruences

$$x \equiv_7 18\,, \quad x \equiv_{12} 3\,, \quad x \equiv_5 7\,, \quad x \equiv_{28} 11\,.$$

In this case $m_1 = 7$, $m_2 = 12$, $m_3 = 5$, $m_4 = 28$ and then m_1, m_4 have common factors, and m_2, m_4 too. We focus on the first and last congruences. As $28 = 4 \times 7$, we replace the last using rule nr. 1:

$$x \equiv_4 7 \quad x \equiv_7 11\,.$$

Then we consider the two congruences modulo 7: $x \equiv_7 18$ and $x \equiv_7 11$. For compatibility we must check that 18 and 11 are congruent modulo 7. This is true, and then we remove one of them. We continue with the congruence modulo 12, that we decompose using rule nr. 1 in:

$$x \equiv_4 3 \quad x \equiv_3 3\,.$$

So we have now two congruences modulo 4 in the system, with right hand side 11 and 3. As 11 and 3 are in the same class modulo 4, we can remove one of them. Next, we write the steps that we carried over the system (the numbers between parentheses refer to the line to which the rule is applied):

$$\begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_{12} 3\,, & (2) \\ x \equiv_5 7\,, & (3) \\ x \equiv_{28} 11\,. & (4) \end{cases} \overset{\text{rule } 1_{(4)}}{\rightarrow} \begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_{12} 3\,, & (2) \\ x \equiv_5 7\,, & (3) \\ x \equiv_7 11\,, & (4) \\ x \equiv_4 11\,. & (5) \end{cases} \overset{\text{rule } 2_{\ (1,4)}}{\rightarrow} \begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_{12} 3\,, & (2) \\ x \equiv_5 7\,, & (3) \\ x \equiv_4 11\,. & (4) \end{cases}$$

$$\overset{\text{rule } 1_{(2)}}{\rightarrow} \begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_4 3\,, & (2) \\ x \equiv_3 3\,, & (3) \\ x \equiv_5 7\,, & (4) \\ x \equiv_4 11\,. & (5) \end{cases} \overset{\text{rule } 2_{(2,5)}}{\rightarrow} \begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_3 3\,, & (2) \\ x \equiv_5 7\,, & (3) \\ x \equiv_4 11\,. & (4) \end{cases}$$

now we can apply the Chinese Remainder Theorem. We can still simplify the system a bit more, with smaller numbers, for example, in (1), 18 is congruent

to 4 modulo 7, in (3) 7 es congruent to 2 modulo 5 and in (4), 11 is congruente to 3 modulo 4. That is:

$$\begin{cases} x \equiv_7 18\,, & (1) \\ x \equiv_3 3\,, & (2) \\ x \equiv_5 7\,, & (3) \\ x \equiv_4 11\,. & (4) \end{cases} \rightarrow \begin{cases} x \equiv_7 4\,, & (1) \\ x \equiv_3 3\,, & (2) \\ x \equiv_5 2\,, & (3) \\ x \equiv_4 3\,. & (4) \end{cases} \overset{\text{rule 1 (2,4)}}{\rightarrow} \begin{cases} x \equiv_7 4\,, & (1) \\ x \equiv_{12} 3\,, & (2) \\ x \equiv_5 2\,, & (3) \end{cases}$$

By applying the CRT: $M = 7 \times 12 \times 5 = 420$.

$$\begin{aligned} 60 \times c_1 &\equiv_7 4 \\ 35 \times c_2 &\equiv_{12} 3 \\ 84 \times c_3 &\equiv_5 2 \end{aligned}$$

$7 \times 8 = 56$ then 60 is 4 modulo 7, and then $c_1 \equiv_7 1$. On the other side 35 is -1 modulo 12, and then $c_2 \equiv_{12} (-3) \equiv_{12} 9$. And last, 84 is -1 modulo 5, so $c_3 \equiv_5 -2 \equiv_5 3$. The solution is given by:

$$x \equiv_{420} 1 \times 60 + 9 \times 35 + 3 \times 84 \equiv_{420} 627 \equiv_{420} 207.$$

Checkup:

$$\begin{aligned} 207 - 18 &= 189 \text{ is divisible by } 7\,, \\ 207 - 3 &= 204 \text{ is divisible by } 12\,, \\ 207 - 7 &= 200 \text{ is divisible by } 5\,, \\ 207 - 11 &= 196 \text{ is divisible by } 28. \end{aligned}$$

5.5. Congruences and powers

In this Section, we prove the so-called *Fermat's Little Theorem* (FLT) from Euler's theorem and we connect the congruences and the powers of integer numbers.

Definition Given a natural number m, the function $\varphi(m)$ is defined as the *amount of integers x such that $0 < x < m$ that are coprime with m.* Such function is called *Euler's function.*

Examples $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(6) = 2$, $\varphi(7) = 6$. In general, if p is prime, $\varphi(p) = p - 1$.

Properties Euler Function

a) $\varphi(m)$ gives the number of integers that are invertible modulo m ($\mathbb{Z}_m^*$).
b) If m and n are coprime, then $\varphi(m \cdot n) = \varphi(m) \cdot \varphi(n)$.
c) If p is prime, then $\varphi(p^n) = p^{n-1}(p-1)$.

Proof a) We already know that $\gcd(a, m) = 1$ if and only if a has an inverse (modulo m). Then the amount of coprimes with m and less than m is the same as the

amount of invertibles in $\mathbb{Z}_m$.
b) We will see that for each a coprime with $m \cdot n$ (so coprime with each of them) and $0 < a < m \cdot n$, there exists a unique b coprime with m such that $0 < b < m$ and a unique c coprime with n, such that $0 < c < m$. Let us consider the integer division of a by m and a by n, and define b and c as the respective remainders:

$$\begin{aligned} a &= q_1 m + b \\ a &= q_2 n + c \end{aligned}$$

We will check that b and c satisfy the desired conditions. Evidently $0 \leq b < m$. If b were zero we would have that m divides a, but we know that a is coprime with m. Then we must have the strict inequality $0 < b < m$. On the other side, if we had a common divisor of b and m, it must be a divisor of a too, and then m and a would have a common divisor. We conclude that b is coprime with m, and so, it is in the list of numbers counted by Euler's function, because it is strictly between zero and a. We can make exactly the same argument with n and c. Then, given a coprime with $m \cdot n$ and $0 < a < m \cdot n$, we showed that there is a pair (b, c) such that b is coprime with m and $0 < b < m$ and c is coprime with n such that $0 < c < n$. Now we will show that if a' is another coprime with $m \cdot n$ such that $0 < a' < m \cdot n$ then it must have a different pair (b', c'). If we had the same pair for a and a' we would have that (starting from b):

$$\begin{aligned} a &= q_1 m + b \\ a' &= h_1 m + b \end{aligned}$$

then

$$a - a' = (q_1 - h_1)\, m \Rightarrow m | (a - a')$$

In the same manner, we conclude that $n|(a - a')$. As m and n are coprime, we have that $m \cdot n|(a - a')$. But $|a - a'| < m \cdot n$, because a and a' are positive and less than $m \cdot n$, then we must have that $|a - a'| = 0$, so $a = a'$. That is, each a, coprime with $m \cdot n$, can be identified with a unique pair (b, c) such that b is coprime with m and c is coprime with n, by using the remainders of the division by m and n respectively.
Now we will show that every pair (b, c) such that $0 < b < m$, b coprime with m and $0 < c < n$ with c coprime with n there corresponds an integer a such that $0 < a < mn$ and coprime with $m \cdot n$. We look for an a satisfying:

$$\begin{aligned} a &= q_1 m + b \\ a &= q_2 n + c \end{aligned}$$

The numbers q_1 and q_2 should satisfy:

$$q_1 m + b = q_2 n + c \Rightarrow q_1 m - q_2 n = c - b\,.$$

As m is coprime with n (and this is very important), Euclid's algorithm guarantees a solution to this problem where the general solution (see the section about linear diophantine equations) is given by:

$$\begin{aligned} q_1 &= x_0 + \gamma n \\ q_2 &= y_0 + \gamma m \end{aligned}$$

such that γ is an arbitrary integer. So, we define the number a by means of the general solution, that is defined up to a multiple of $m \cdot n$:

$$a = (x_0 + \gamma n)\, m + b = (y_0 + \gamma m)\, n + c$$

we can choose γ such that $0 \leq a < m \cdot n$. We must now check that the a chosen in this way is coprime with $m \cdot n$. If they had a common divisor d, then d would divide a and one of m or n. If it divides m, then it must divide b, and then m and b would not be coprime.
Concluding, we showed that there is *bijective* map between the numbers a that are coprime with $m \cdot n$ and the set of pairs (b, c), such that b is coprime with m and c coprime with n and $0 < b < m$ and $0 < c < n$. Then, they must have the same number of elements. The pairs (b, c) have $\varphi(m)\varphi(n)$ elements and this number must be equal to the amount of numbers a coprime with $m \cdot n$, such that $0 < a < m \cdot n$ that is defined as $\varphi(m \cdot n)$
c) If a is not coprime with p^n, it must have at least one factor p in its factorization. That is, it must be of the form $a = kp$, so if $0 < a < p^n$ then $0 < k < p^{n-1}$. Then, there are exactly p^{n-1} that are not coprime with p^n, so there must be $p^n - p^{n-1}$ that are coprime with p^n and strictly smaller than p^n. ■

Note If a number m is factorized as a product of primes, then, by using properties b) and c):

$$m = \prod_{i=1}^{N} p_i^{n_i} \Rightarrow \varphi(m) = \prod_{i=1}^{N} p_i^{n_i - 1}\,(p_i - 1)\cdot$$

that can be written as:

$$\varphi(m) = \prod_{i=1}^{N} p_i^{n_i}\,(1 - 1/p_i) = m \prod_{i=1}^{N} (1 - 1/p_i)\ .$$

The number obtained is an integer, as we can check that $p_i^{n_i}\,(1 - 1/p_i) = p_i^{n_i - 1}\,(p_i - 1)$.

Theorem Euler-Fermat If a and m are coprime, then

$$a^{\varphi(m)} \equiv_m 1\,.$$

Proof Let $k := \varphi(m)$, and let $m_1, \ldots, m_k$ be the integers between 1 and m that are coprime with m. Consider now the numbers

$$am_1\,, \ldots\,, am_k\,.$$

We will show that am_i must be congruent modulo m with some m_j. We take the integer division:

$$am_i = qm + r\,,$$

so that $0 < r < m$ (it cannot be zero because m is coprime with am_i). The number r must be coprime with m, because if d divides r and m then it divides am_i, so d would be a common divisor of m and am_i. So, r is coprime with m and it must be some of the m_j in the list. On the other hand, am_i cannot be in the same class modulo m with am_j for $i \neq j$. In that case we would have that

$$am_i - am_j = qm \Rightarrow a\,(m_i - m_j) = qm\,.$$

Being m coprime with a, m would divide $m_i - m_j$. But as $|m_i - m_j| < m$ this is not possible. We conclude that $am_i \equiv_m m_j$ for some j, and $am_i \not\equiv_m am_j$ if $i \neq j$. Then, if we multiply the numbers am_i, the product is congruent modulo m with the product of the m_i:

$$(am_1)(am_2)\ldots(am_k) = a^k m_1 \ldots m_k \equiv_m m_1 \ldots m_k\,.$$

That is, as $m_1 \ldots m_k$ are coprime with m, they have an inverse (mod m) and then:

$$a^k \equiv_m 1\,.$$

■

Corollary FERMAT'S LITTLE THEOREM If p is prime and a is not divisible by p, then we must have that

$$a^{p-1} \equiv_p 1\,.$$

Remark-1: By multiplying both sides by a we obtain the relationship

$$a^p - a \equiv_p 0\,,$$

that implies Fermat's little theorem if p does not divide a, because in that case a has an inverse modulo p.

Remark-2: The previous remark is valid for p prime and *for every integer* a. If p divides a the result is trivial, because in that case both a^p and a are congruent with 0.

Corollary Let p be a prime. Then, for every integer a such that $p \nmid a$ and for every positive integer n, we have that

$$n \equiv_{p-1} r \Longrightarrow a^n \equiv_p a^r\,.$$

Proof As n is congruent with r modulo $p-1$, we have that $n = q\,(p-1) + r$, then

$$a^n = \left(a^{(p-1)}\right)^q a^r \underbrace{\equiv_p}_{\text{FLT}} 1 \times a^r\,.$$

■

Notice that r can be any number that is congruent with n modulo $p-1$, in particular, the remainder of dividing n by $p-1$.

Example-1 Compute the remainder of the division between 12^{2013} and 7.

$2013 = q \times 6 + r$, with $r = 3$ and 12 is congruent with 5 modulo 7:

$$12^{2013} \equiv_7 5^{2013} \equiv_7 5^{q\times 6+3} \equiv_7 \underbrace{\left(5^6\right)^q}_{\equiv_7 1} 5^3 = 125\,.$$

The remainder is 6 ($125 = 17 \times 7 + 6$).

Example-2 Compute the remainder of the division between $16^{12^{2013}}$and 5.

By the last corollary, if $a = 16$, and $n = 12^{2013}$, $a^n \equiv_5 a^r$ where r is congruent with n modulo 4. Then:

$$r \equiv_4 12^{2013} \equiv_4 0^{2013} = 0\,.$$

so then

$$a^n \equiv_5 1\,,$$

and the remainder is 1.

When the divisor is not prime

If we want to find the remainder of a^n divided by m, and the number m is not prime, then it can be factored as $p_1 \dots p_k$. Applying the corollary to compute the remainder of the division of a^n by p_i and combining with the CRT (Chinese Remainder Theorem) we can obtain a solution. We compute the remainders $r_i :=$ remainder of the division between a^n by p_i, by means of direct inspection or by the Fermat's little Theorem, so we find the following system of congruences:

$$x \equiv_{p_1} r_1\,, x \equiv_{p_2} r_2\,, \dots, x \equiv_{p_k} r_k \iff x \equiv_m x_0 \quad (m = p_1 p_2 \dots p_k)$$

where x_0 is a solution found by the CRT.

Example Compute the remainder of the division of $3^{2^{25}}$ by 390.

$390 = p_1 \times p_2 \times p_3 \times p_4 = 2 \times 3 \times 5 \times 13$. We compute first the remainders:

$$3^{2^{25}} \equiv_2 1^{2^{25}} = 1 = r_1\,,$$

$$3^{2^{25}} \equiv_3 0^{2^{25}} = 0 = r_2\,,$$

$$3^{2^{25}} \equiv_5 3^0 = 1 = r_3\,,\ \text{because } 2^{25} = 2 \times 2^{24} \equiv_4 0\,,$$

by applying Fermat's little theorem (FLT) with $p = 5$. To find r_4, that is the remainder of dividing $3^{2^{25}}$ by 13, we may apply FLT by computing the remainder of the division of 2^{25} by 12 . In this case is better to write $12 = 3 \times 4$. We compute

$$2^{25} \equiv_3 (-1)^{25} = -1 \equiv_3 2\,, \quad 2^{25} \equiv_4 0\,.$$

Now we have that

$$r \equiv_3 2 \qquad r \equiv_4 0\,.$$

Then r must be a multiple of 4, that is $r = k4$ and

$$4k \equiv_3 2 \Rightarrow k \equiv_3 2\,,$$

where we used that 4 equal to 1 (modulo 3) thus obtaining:

$$r = 4k = 8\,.$$

Then, $r_4 \equiv_{13} 3^8 = 3^3 \times 3^3 \times 3^2 \equiv_{13} 9$, because $3^3 = 27$ is congruent with 1 modulo 13.

6. Groups

Bibliography:

Algebra for Computer Science, Lars Gårding, Torbjörn Tambour.

Abstract Algebra: Theory and Applications, Thomas Judson.

6.1. Definitions and general properties

A group is a non-empty set $G = \{a, b, c, ...\}$ with an operation $(a, b) \mapsto ab$ from $G \times G$ to G such that:

(i) $(ab)c = a(bc)$ for all $a, b, c \in G$ (associativity),

(ii) G has a *unit element* (identity or neutral element) e such that $ae = ea = a$ for all $a \in G$,

(iii) Every $a \in G$ has an *inverse*, that is, there exists an element $b \in G$ such that $ba = ab = e$. The inverse of a is written a^{-1}.

If the operation is conmutative, i.e. $ab = ba$ for all $a, b \in G$, the group is called conmutative or *abelian*.

Proposition In a group, there is only one neutral element, and the inverse of an element is unique.

Proof Assume that e and e' are two neutral elements. In that case:

$$e = ee' = e' .$$

Assume now that an element a has two inverses b, b'.

$$b = be = bab' = eb' = b' .$$

■

Examples The sets defined as the integer multiples of a fixed number a, ie $\mathbb{Z}a$, are a group with respect to the 'sum' operation. The set of rational numbers $q \in \mathbb{Q}$ greater than zero and the set of real numbers $x \in \mathbb{R}$ greater than zero, form respective groups with multiplication.

Definition A *subgroup* of a group G is a non empty subset $H \subset G$ that is itself a group under the same operation as G.

Every group G has some natural subgroups, as for example its *center* $\text{cent}(G)$ that is given by all the elements $a \in G$ that conmute with every other element $b \in G$,

that is $ab = ba$ (if G is abelian then cent $(G) = G$). For every $a \in G$ there is a group called *centralizer* $C(a)$ that is formed by all the elements b that conmute with a. The center is always an abelian subgroup, but the centralizer does not need to.

Lemma Let H be a subgroup of G. Let $\sim$ be the relationship between elements of G defined by the rule

$$a \sim b \Leftrightarrow ab^{-1} \in H\,.$$

Then, $\sim$ is an equivalence relationship.

Proof An equivalence relationship is i) reflexive, that is $a \sim a$. In this case we must show that $aa^{-1} \in H$. This is true because $aa^{-1} = e \in H$ (H is a subgroup and so contains the neutral element). ii) The relationship must be symmetric, that is $a \sim b \Rightarrow b \sim a$. Then, we must show that if $ab^{-1} \in H$ then $ba^{-1} \in H$. This is true because ba^{-1} is the inverse of ab^{-1} and H contains the inverses of all its elements. iii) transitivity: given a, b, $c \in G$ such that $a \sim b$ and $b \sim c$, we must show that $a \sim c$. That is, if $ab^{-1} \in H$ and $bc^{-1} \in H$ then $(ab^{-1})(bc^{-1}) = ac^{-1} \in H$ ■

Definition Given $\sim$ an equivalence relationship on a set X and $a \in X$. The *equivalence class* of a is

$$[a] = \{x \in X | x \sim a\}\,.$$

Definition Given a set X, a partition $\mathcal{P}$ of X is a collection of subsets A_i, $i \in I$, such that **(1)** A_i *cover* X, that is, $\cup_{i\in I} A_i = X$. **(2)** A_i are pairwise disjoint, that is, if $i \neq j$ then $A_i \cap A_j = \emptyset$.

Lemma An equivalence relationship $\sim$ over X defines a unique partition of X. The sets of the partition are the equivalence classes of $\sim$. Conversely, any partition of X defines an equivalence relationship. That is, given a partition $\mathcal{P}$ of X we can construct an equivalence relationship over X such that the associated classes are given by $\mathcal{P}$.

Proof Exercise.

Example The *classes modulo* m define a partition of $\mathbb{Z}$ that corresponds to the equivalence relationship: $a \sim b$ if and only if $m|(a - b)$.

Generators Let $g = ab \ldots c$ be a composition of elements of G. The inverse of g, g^{-1}, is given by the expression

$$c^{-1} \ldots b^{-1}a^{-1}.$$

To verify this is enough to compose this element with g and apply the associative property. This shows that every non-empty set M of a group G generates a subgroup given by the set of compositions $a_1a_2...a_n$ such that a_i or a_i^{-1} are in M.

Cyclic Groups Any element a of a group G generates what is called a *cyclic group* given by the set

$$\{a^n : n \in \mathbb{Z}\} \qquad \left(a^0 = e\right)\,.$$

of all the powers, positive, negative or zero, of the element a. The power of exponent zero corresponds to the unit element. A cyclic group may be infinite, in which case

each power a^n is a different element. For a proof, assume that for different exponents we have $a^n = a^m$ with $m < n$, then $a^{n-m} = e$ and $n - m > 0$. In this case, consider the least integer $r > 0$ such that $a^r = e$. The group would have exactly r elements given by

$$\left\{a^k : 0 \leq k < r\right\} .$$

The number of the elements in the finite group is called the *order* of the group.

DIRECT PRODUCT If G and H are groups, the cartesian product $G \times H$ has a group structure. Its elements are pairs (a, b) with $a \in G$, $b \in H$ and the operation is defined as:

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2, b_1 b_2) .$$

6.2. Permutation Groups

Let $X = \{x_1, x_2, \dots\}$ be a set. All the bijections $f : X \to X$ of X in itself are a group under composition.

$$(fg)(x) := f(g(x)) .$$

The neutral element is the identity function, sending each element x to itself: $x \to x$ and the inverse of f is the function that sends each y to the value x such that $y = f(x)$ (this value exists because f is bijective).

If X has a finite amount of elements $x_1, \dots, x_n$, a bijection corresponds to a *permutation* of its elements. That is, the elements

$$f(x_1), f(x_2) \dots f(x_n)$$

are the same as $x_1, \dots x_n$ but possibly in a different order.

The group of all bijections of a set is called the group of permutations. Subgroups of such groups are universal examples of groups. In fact, every group is essentially a subgroup of a group of permutations conveniently defined.

The following conditions define subgroups in a group of bijections:

i) Every f that fixes an element $x \in X$ (this is the *stabilizer* of x).

ii) Every f that leaves invariant a set $Y \subset X$, i.e. $f(Y) = Y$.

6.3. Group morphisms and quotient groups

We will assume that the groups are abelian.

Definition MORPHISMS An application $f : G \to H$ where G and H are groups is called a *group morphism* or *homomorphism* if

$$f(ab) = f(a)f(b) \quad \forall a, b \in G .$$

(Note that the composition of the right hand side is in the group H, but the one on the left hand side is in G.) An *isomorphism* is a bijective homomorphism. Note that the definition implies that $f(e_G) = e_H$, because $f(a) = f(ae_G) = f(a)f(e_G)$, and then $f(e_G)$ is the neutral element of H and must be equal to e_H (due to uniqueness), ie.: $f(e_G) = e_H$.

Example The application $x \to e^x$ is an isomorphism from $(\mathbb{R}, +)$ to $(\mathbb{R}^+, \cdot)$ (that is, from real numbers with addition to positive reals with multiplication).

Properties Let $f : G \to H$ be a morphism between groups.

i) The *image* of f, $\mathrm{Im}(f) := f(G)$ is a subgroup of H.

ii) The *kernel* of f, $\mathrm{Ker}(f) := f^{-1}(e)$ (Where e is the neutral element of H), is a subgroup of G.

Proof

i) given h_1 and h_2 in $\mathrm{Im}(G)$ we will check that $h_1 h_2^{-1}$ also belongs to $\mathrm{Im}(G)$. As these elements belong to the image of f, there must exist g_1, g_2 in G such that $f(g_1) = h_1$ and $f(g_2) = h_2$. Given $g = g_1 g_2^{-1}$, then $f(g) = f(g_1)f(g_2^{-1}) = h_1 h_2^{-1}$ (a homomorphism transforms inverses in inverses). Then, $h_1 h_2^{-1}$ belongs to the image of f.

ii) Let g_1 and g_2 be in $\mathrm{Ker}(f)$. Clearly, $g_1 g_2^{-1}$ is in the kernel too, because $f(g_1 g_2^{-1}) = f(g_1)f(g_2^{-1}) = e \cdot e = e$. Moreover, if $x \in \mathrm{Ker}(f)$, then for all $g \in$ G:

$$f(gxg^{-1}) = f(g)ef(g^{-1}) = f(g)f(g^{-1}) = e$$

and then gxg^{-1} belong to $\mathrm{Ker}(f)$ for every $g \in G$. ■

Theorem Let G be a group (abelian) and H a subgroup. Then, the quotient classes of H in G form a group, denoted by G/H, that is the so-called *quotient group of* G modulo H. The composition rule in G/H is given by

$$(aH)(bH) = abH.$$

Moreover, there is an homomorphism $\phi : G \to G/H$, defined as $\phi(g) = gH$ and the kernel of ϕ is H.

Proof: Remember that the elements of the quotient group are *classes* such that for each element a, b belongs to the same class if and only if

$$b \in aH \Rightarrow [a] = aH\,.$$

This means that $b = ah$ for some $h \in H$. It is easy to show that the class $eH = H$ is the neutral element of G/H and the inverse of aH is $a^{-1}H$. Moreover, the kernel of ϕ are the elements $x \in G$ such that $\phi(x) = H$, that is, they are the $x \in G$ such that $xh \in H$ for every $h \in H$. Then, if $xh = h'$ we have that $x = h'h^{-1} \in H$, and then $x \in H$. ■

Theorem Group morphisms Let $f : G \to H$ be a group morphism. Then there exists an isomorphism between $G/\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$.

Proof We define a function $\bar{f} : G/\mathrm{Ker}(f) \to \mathrm{Im}(H)$ by means of the expression:

$$\bar{f}(x\,\mathrm{Ker}(f)) = f(x)\,.$$

We first show that the function is well-defined. If $x\,\mathrm{Ker}(f) = y\,\mathrm{Ker}(f)$ (x and y are representatives of the same class) then $y^{-1}x \in \mathrm{Ker}(f)$ and

$$f(y^{-1}x) = e \Rightarrow f(x) = f(y)\,.$$

It is injective because if $\bar{f}(x\,\mathrm{Ker}(f)) = e$ then $f(x) = e$ so $x \in \mathrm{Ker}(f)$ and then $x\,\mathrm{Ker}(f) = \mathrm{Ker}(f)$. Let us see that it is surjective. Given $h \in \mathrm{Im}(f)$, we look for $f^{-1}(h)$. Given any $x \in G$ such that $f(x) = h$ (x exists because h is in the image of f), we have that

$$\bar{f}(x\,\mathrm{Ker}(f)) = h\,.$$

■

Note: if f is injective, $\mathrm{Ker}(f) = \{e\}$ then $x\,\mathrm{Ker}(f)$ has only one element that can be identified with x. In this case $\bar{f}$ defines an isomorphism from G to $\mathrm{Im}(H)$.

6.4. Finite Groups

Definitions The total number of elements of a set X is called its *order* and is denoted by $|X|$. The *order* of an element a of a group G is the order n of the cyclic subgroup generated by a. This number n is also know as the *period,* because every element of the sequence of powers of a show up after n steps. The number of classes of a subgroup H of a group G is called the *index* of H in G and is denoted by $[G : H]$.

Lemma Let G be a finite group and H a non-empty subset, closed under the operation. Then, H is a subgroup of G.

Proof It is enough to prove that the inverses of elements in H belong to H, because it is already closed under the group operation. Let $a \in H$. If $a = e$ then $a^{-1} = e$ and so it is in H. Then we assume that $a \neq e$. Consider the powers of a:

$$a, a^2, a^3, ...$$

All these elements are in H due to its closedness under the group operation. Being H finite, this sequence is a finite set, so we must have repetitions. Then, for some pair m, n we must have:

$$a^m = a^n.$$

Assuming that $m < n$ we multiply both sides by $a^{-m} (= (a^{-1})^m)$ and we have that

$$e = a^{n-m}\,.$$

Being $a \neq e$, then $n - m > 1$ and we can write

$$e = a \cdot a^{n-m-1} \Rightarrow a^{-1} = a^{n-m-1}.$$

So H is closed under the inverse and then it is a subgroup of G ■

Proposition Let H be a subgroup of G with $g \in G$. The application $\phi : H \to gH$ defined by $\phi(h) = gh$, is bijective, that is, the number of elements in H is the same as the number of elements in gH.

Proof: Let us see that the application is injective. If $\phi(h_1) = \phi(h_2)$ then $gh_1 = gh_2$, cancelling both g we have that $h_1 = h_2$. It is obviously surjective (given $x = gh$, then $\phi(h) = x$), then it is bijective. ■

Lagrange's Theorem

Quotient classes form a partition of the Group G, we have the following result, connecting the theory of finite groups with number theory.

Theorem (LAGRANGE 1775). Let G be a finite group and let H be a subgroup of G. Then, $|G| = [G : H] \times |H|$. In other words, the order of H divides the order of G.

Proof The group G is partitioned in $[G : H]$ different classes. Each class has $|H|$ elements; then, $|G| = [G : H]|H|$.■

Corollary If G is a finite group and $g \in G$, then the order of g divides the order of G.

Proof The order of g is the number of elements in the cyclic group generated by g ■

Corollary Let $|G| = p$ with p prime. Then G es cyclic and any $g \in G$ such that $g \neq e$ is a generator.

Proof The order of the cyclic group generated by g cannot be 1 because $g \neq e$. This number must divide p, so it must be equal to p and then generates all the group. This is valid for every member of G that is not the neutral element ■

This last corollary suggests that the groups of prime order p must be similar to $\mathbb{Z}_p$

Corollary Let H and K be subgroups of finite order G such that $G \supset H \supset K$. Then

$$[G : K] = [G : H][H : K].$$

Proof Observe that $[G : K] = \frac{|G|}{|K|} = \frac{|G|}{|H|}\frac{|H|}{|K|} = [G : H][H : K]$.■

Euler-Fermat's theorem from Lagrange's Theorem

Proposition Let $k \in \mathbb{Z}$, then we know that $k \neq 0$ has a multiplicative inverse in $\mathbb{Z}_n$ if and only if k is relatively prime to n. We denote the set of these $k \in \mathbb{Z}_n$ by $\mathbb{Z}_n^*$. Then, $\mathbb{Z}_n^*$ has a group structure, called the group of *units* of $\mathbb{Z}_n$.

Proof To show that $\mathbb{Z}_n^*$ is a group, we show that: 1) $1 \in \mathbb{Z}_n^*$, true; 2) Every element in $\mathbb{Z}_n^*$ has inverse: by definition; 3) it is closed under multiplication: if a and b belong to $\mathbb{Z}_n^*$, then a and n are coprime and b and n are coprime if and only if ab and n are coprime. By Bézout identity, there is an inverse of ab modulo n ■

This implies that the order of the group $\mathbb{Z}_n^*$ is the same as the value of the Euler function evaluated in n: $\varphi(n)$. That is, $|\mathbb{Z}_n^*| = \varphi(n)$.

Theorem (Euler) Let a and n be integers such that $n > 0$ and $\gcd(a, n) = 1$. Then $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Proof We know that the order of $\mathbb{Z}_n^*$ is $\varphi(n)$. The order d of the cyclic subgroup generated by a must divide $\varphi(n)$, that is $\varphi(n) = qd$ and $a^d \equiv_n 1$. Then, $a^{\varphi(n)} = \left(a^d\right)^q \equiv_n 1$. ■

If we consider the special case of Euler's Theorem in which $n = p$ prime and recall that $\varphi(p) = p - 1$, we obtain the so-called *Fermat's Little Theorem.*

Theorem (Fermat's Little Theorem) Let p be any prime number and assume that $p \not| a$. Then, $a^{p-1} \equiv_p 1$. Moreover, for any integer b, $b^p \equiv_p b$.

Proof If p does not divide a, we apply Euler's Theorem. For the second assertion, if p divides b the identity is trivial because $b^p - b$ is divisible by p. If p does not divide b then we apply the first part, ie., $b^{p-1} \equiv_p 1$ and multiply by b both sides ■

Homomorphisms of finite groups

For some applications we need properties of the homomorphisms between finite order groups. Some of them are direct applications of the former theorems. For example:

Properties: The kernel K of a homomorphism $\phi : G \to H$ is a subgroup of G (this is a general theorem valid for any group homomorphism), then its order, by Lagrange Theorem, must divide the order of G.
If G is cyclic, K is also cyclic (general theorem), and then, there must exist in G an element of the same order of K.

Property: If a is in the image of ϕ and has order m, then G has an element whose order is divisible by m.

Proof: If $a \in \mathrm{Im}\,(\phi)$, we consider $x \in G$ such that $a = \phi(x)$. If n is the order of x, $\phi(x^n = e_G) = \phi(x)^n = a^n = e_H$, then m divides n ■

Direct products and isomorphisms

We summary two characterisations of finite groups:

If the order of a group G is p, a prime, then the group has a generator, call it g, such that

$$x \in G \Rightarrow \exists\, n, 0 \leq n < p : x = g^n\,,$$

that is, a group of order prime is cyclic and we can define the following isomorphism from G to $\mathbb{Z}_p$ (with the sum operation):

$$\phi(x) = n \quad (x = g^n) \ .$$

n is the unique value with this property because p is the order of the group and $g \neq e$ is its generator.

Another characterization already seen is that, if m and n are relative primes, then

$$\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$$

where $\cong$ means that there is an isomorphism between both groups. In this case, given $[x] \in \mathbb{Z}_{mn}$, we define the isomorphism as:

$$\phi([x]_{mn}) = ([x]_m, [x]_n) \ ,$$

The function is well defined, because if x and y are in the same class modulo mn, then we can see that $\phi([x]_{mn}) = \phi([y]_{mn})$. If

$$[x]_{mn} = [y]_{mn}$$

we must have that

$$x - y = qmn$$

then

$$[x]_m = [y]_m \text{ and } [x]_n = [y]_n$$

because the difference is divisible by m and by n. So the mapping is well defined, because it does not depend of the representative element selected.

On the other side, if we take two numbers, x and y such that:

$$x \equiv_m y \, , \quad x \equiv_n y$$

then

$$x - y \equiv_m 0 \text{ and } x - y \equiv_n 0$$

Then $x - y$ is divisible by m and by n so $x - y$ must be divisible by mn because m and n are relative primes. So, the mapping is *injective*. As the amount of elements of both sets are the same (mn) then it is also *surjective*.■

Now we will study in more detail the structure of the product groups. Given two groups G and H, it is possible to create a new group with the cartesian product of G and H, $G \times H$. Reciprocally, given a group, it is sometimes possible to decompose it, in the sense that there is an isomorphism between the given group and a direct product of "smaller" groups. Instead of studying a group G is, in general, easier to study its components.

If $(G, \cdot)$ and $(H, *)$ are groups, we can give a group structure to $G \times H$, that is, to the set of pairs (g, h) with $g \in G$ and $h \in H$. We may define a binary operation on $G \times H$ as $(g_1, h_1)(g_2, h_2) = (g_1 \cdot g_2, h_1 * h_2)$; in other words, we apply the operation of each group in each slot. In what follows we will omit the symbols of the operations writing only $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$.

Proposition Let G and H be groups. The set $G \times H$ is a group under the operation

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$$

where g_1, $g_2 \in G$ y h_1, $h_2 \in H$.

Proof Clearly the operation is well defined and it is closed. If e_G and e_H are the identities of the groups G and H respectively, then (e_G, e_H) is the identity in $G \times H$. The inverse of $(g, h) \in G \times H$ is (g^{-1}, h^{-1}). The associativity is obtained directly from associativity in G and in H ■

Definition The group $G \times H$ is the so-called *external direct product* of G and H. The same definition may be extended to more than two groups. The direct product

$$\Pi_{i=1}^{n} G_i = G_1 \times G_2 \times \cdots \times G_n$$

of groups G_1, G_2, $\ldots G_n$ is defined analogously (operations are defined for each slot). If the groups are identical, that is if $G = G_1 = G_2 = \cdots = G_n$, we write G^n instead of $G_1 \times G_2 \times \cdots \times G_n$.

Theorem Let $(g, h) \in G \times H$. If g and h have finite orders r and s respectively, then the order of (g, h) in $G \times H$ is the least common multiple of r and s.

Proof Let m be the least common multiple of r and s and let $n = |(g, h)|$ (ie, the order of the element (g, h)). We have that $(g, h)^m = (g^m, h^m) = (e_G, e_H)$. Then, n must divide m and $n \leq m$. On the other hand, r and s must divide n; then n is a multiple of r and of s. Being m the least common multiple of r and s, $m \leq n$. Then, we must have that $m = n$ ■

Corollary Let $(g_1, \ldots, g_n) \in \Pi_{i=1}^{n} G_i$. If g_i has finite order r_i in G_i, Then the order of $(g_1, \ldots, g_n)$ in $\Pi_{i=1}^{n} G_i$ is the least common multiple of r_1, $\ldots$, r_n.

The following Theorem tells us exactly when a direct product of cyclic groups is itself cyclic.

Theorem The group $\mathbb{Z}_m \times \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_{mn}$ if and only if $\gcd(m, n) = 1$.

Proof We already showed the implication $\Leftarrow$, that is, if m, n are relative primes there is an isomorphism between both groups. Assume now that $\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$, and we will show that $\gcd(m, n) = 1$. We will show that if $\gcd(m, n) = d > 1$, then $\mathbb{Z}_m \times \mathbb{Z}_n$ cannot be cyclic (and then there cannot be an isomorphism between both groups). Notice that mn/d is divisible by both m and n and if $d > 1$ mn/d is *strictly smaller* than mn; then, for any element $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_n$

$$\underbrace{(a, b) + \cdots + (a, b)}_{mn/d \text{ times}} = (0, 0)$$

then there is no (a, b) that generates $\mathbb{Z}_m \times \mathbb{Z}_n$, because it must generate mn elements, and it generates only mn/d ■

Corollary Let $n_1, \ldots, n_k$ be positive integers. Then

$$\prod_{i=1}^{k} \mathbb{Z}_{n_i} \cong \mathbb{Z}_{n_1 \ldots n_k}$$

if and only if $\gcd(n_i, n_j) = 1$ for $i \neq j$.

Corollary If $m = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$, where the p_i are different primes, then

$$\mathbb{Z}_m \cong \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_k^{e_k}} .$$

Proof Using the fact that $p_i^{e_i}$ are relative primes by pairs, the result is obtained from the previous corollary ■

7. Rings and Fields

Bibliography: *Abstract Algebra: Theory and Applications,* Thomas Judson. Free download: http://abstract.ups.edu/download.

7.1. Definitions and general properties

A non-empty set A is a *ring* if it posess two binary operations "addition" and "multiplication", satisfying the following properties:

1. $a + b = b + a$ for all $a, b \in A$. (Conmutativity of addition)

2. $(a + b) + c = a + (b + c)$ for all $a, b, c \in A$. (Associativity of addition)

3. There exists an element $0_A \in A$ (Neutral element for addition) such that $a+0_A = a$ for all $a \in A$.

4. For every element $a \in A$, there exists an element $-a \in A$ such that $a+(-a) = 0_A$. (Existence of inverse w.r.t addition)

5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for $a, b, c \in A$. (Associativity of multiplication)

6. If $a, b, c \in A$, $a \cdot (b + c) = a \cdot b + a \cdot c$, $(a + b) \cdot c = a \cdot c + b \cdot c$. (Right and left distributivity of multiplication w.r.t addition)

The first four properties show that A is an abelian group with respect to the addition operation.

Definition (Ring Homomorphism) If A, A' are rings, a mapping $\phi : A \to A'$ is a *ring homomorphism* if:
(i) $\phi(a + b) = \phi(a) + \phi(b)$ (is a group homomorphism with the addition)
(ii) $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$.
Notice that in these identities the operation on the left is the one of the ring A, and the operation of the right is the one of the ring A'.
If the homomorphism is bijective, it is called a *ring isomorphism.*

7.2. Ring classification

Definitions If there is a unit element $1_A \in A$ such that $1_A \neq 0_A$ and $1_A \cdot a = a \cdot 1_A = a$ for all $a \in A$, we say that A is a ring *with unity* or *identity.* A ring A such that $ab = ba$ for all a, b in A is a *conmutative ring.* A *division ring* is a ring A with unity such that every non-zero element has a multiplicative inverse;

that is, for every $a \in A$ with $a \neq 0_A$, there exists an element a^{-1} such that $a^{-1} \cdot a = a \cdot a^{-1} = 1_A$. A conmutative division ring is a *field*. A commutative ring A with identity is called an *integral domain* if, for every $a, b \in A$ such that $ab = 0_A$, either $a = 0_A$ or $b = 0_A$.

Definition An element $a \neq 0_A$ of a ring A is called a *divisor of zero* if there exists an element $b \in A$, such that $b \neq 0_A$ and $ab = 0_A$. (For example, in $\mathbb{Z}_{12}$, 3 and 4 are divisors of zero).

Proposition Let A be a ring and a, $b \in A$. Then:

1. $a \cdot 0_A = 0_A \cdot a = 0_A$;
2. $a \cdot (-b) = (-a) \cdot b = -a \cdot b$;
3. $(-a) \cdot (-b) = a \cdot b$.

Proof 1. For all a, $a \cdot 0_A = a \cdot (0_A + 0_A) = a \cdot 0_A + a \cdot 0_A$, where we obtain (after cancelling $a \cdot 0_A$ on both sides) $a \cdot 0_A = 0_A$. For $0_A \cdot a$ the proof is similar.

2. Using item 1, $0_A = a \cdot (b - b) = a \cdot b + a \cdot (-b)$, so that the opposite of $a \cdot b$ is $a \cdot (-b)$. The same applies for $(-a)\, b$.

3. Using item 2, $(-a) \cdot (-b) = -\left(a \cdot (-b)\right) = -(-a \cdot b) = a \cdot b$.■

Definition A *subring* S of a ring A is a subset $S \subset A$ such that S is also a ring under the operations inherited from A.

Proposition Let A be a ring and S a subset of A. Then, S is a subring of A if and only if the following conditions are satisfied:

1. $S \neq \emptyset$.
2. $ab \in S$ for every $a, b \in S$ (closed with respect to multiplication).
3. $a - b \in S$ for every a, b in S (subgroup with respect to additon).

Definition An *ideal* of a ring A is a subring I such that if $i \in I$ and $a \in A$, then ia and ai belong to I; that is, $aI \subset I$ and $Ia \subset I$ for all $a \in A$. Every ring has two trivial ideals given by $\{0_A\}$ (the set whose only element is the zero of A) and A itself.

Remark Let A be a ring with unity and I an ideal of A such that $1 \in I$. Then, for every $a \in A$, we must have that $a1 = a \in I$, so $I = A$. In other words, any ideal containing the identity element, must be equal to the whole ring ■

Definition If A is a conmutative ring with unity and $x \in A$, the ideal defined by:

$$\langle x \rangle = \{ax : a \in A\}$$

is called a *principal ideal.*

Proposition Every ideal of the integers $\mathbb{Z}$ is a principal ideal.

Proof Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$ the proposition is proved. If not, let us consider the least natural number in I, and call it c. We will show that $I = \langle c \rangle$. If there exists $z \in \mathbb{Z}$ such that $z \in I$, but $z \notin \langle c \rangle$ then z is not a multiple of c, and then we can write $z = qc + r$ with $0 < r < c$. But in this case, $z - qc = r \in I$ (I is a subring) and we would have a number r such that $0 < r < n$ and $r \in I$. We assumed that n was the least natural number in I. Then, $z \in \langle c \rangle$ and $\langle c \rangle = I$ ■

Proposition The kernel ($\mathrm{Ker}(f)$) of a ring homomorphism $f : A \to B$ is an ideal in A.

Remark When we ask that $aI \subset I$ in the definition of an ideal, we are speaking of *left ideals*, if we ask that $Ia \subset I$, these are called *right ideals.* When we ask both conditions, we are speaking about *bilateral ideals.* If we have a conmutative ring, ideals are always bilateral. Here, we only deal with bilateral ideals.

Theorem Let I be an ideal of A. The quotient group (with respect to addition) A/I is a ring, with multiplication defined as:

$$(a + I) \cdot (b + I) = ab + I\,.$$

Proof We already know that A/I is an abelian group under addition. Let $a + I$ and $b + I$ be in A/I. We must show that the product defined does not depend on the representative element; that is, if $a' \in a + I$ and $b' \in b + I$, then we have to prove that $a'b'$ belongs to $ab + I$. As $a' \in a + I$, $b' \in b + I$, there must be an element $i \in I$ and another $h \in I$ such that $a' = a + i$ and $b' = b + h$. Notice that $a' \cdot b' = (a + i) \cdot (b + h) = a \cdot b + i \cdot b + a \cdot h + i \cdot h$ and that $i \cdot b + a \cdot h + i \cdot h \in I$ because I is an ideal; then $a' \cdot b' \in a \cdot b + I$. Associativity and distributivity are easily verified. ■

Theorem (Canonical Homomorphism) Let I be an ideal of A. The mapping

$$\phi : A \to A/I \quad \text{defined by } \phi(a) = a + I$$

is a ring homomorphism from A over (that is, it is surjective) A/I with kernel I.

Proof The proof of being a surjective group homomorphism is the same as for groups. It only remains to prove that it behaves properly with respect to product. Let a, b be in A:

$$\phi(ab) = (ab + I) = (a + I) \cdot (b + I) = \phi(a) \cdot \phi(b)$$

becase I is an ideal ■

Ejemplo In $\mathbb{Z}$, all the ideals are of the form $\langle n \rangle = \{nz : z \in \mathbb{Z}\} = n\mathbb{Z}$. In this case, the canonical homomorphism is the mapping that for each n it sends an element z to its class in $\mathbb{Z}_n$, that we denoted by $[z]_n$.

7.3. Integral Domains and Fields

Definitions A nonzero element $a \in A$ is is a *zero divisor* if there exists another element $b \in A$ such that $b \neq 0_A$ and $ab = 0_A$. If A is a conmutative ring with unity, is is an *integral domain* if it has no zero divisors. If an element a of a ring A with unity has multiplicative inverse, it is called a *unit.* If every nonzero element of a ring A is a unit, then A is a *division ring.* If it is conmutative too, it is a *Field.*

Examples $\mathbb{Z}_n$ is a commutative ring with unity. If n is not prime, we have elements of $\mathbb{Z}_n$, say q and d such that $qd = n$, and then $qd \equiv_n 0$. That is, if n is not prime, $\mathbb{Z}_n$ has zero divisors. If n is prime, every nonzero element (ie, the numbers not congruent with zero) are invertible (by Bézout identity) and then $\mathbb{Z}_n$ is a field for n prime.

We have the following characterization of integral domains:

Proposition (Cancellation Law) Let A be a conmutative ring with identity. Then A is an integral domain if and only if for every nonzero element $a \in A$ such that $ab = ac$ then $b = c$.

Proof Being A an integral domain, then $ab = ac$ implies that $ab - ac = 0_A$ and then $a\,(b - c) = 0_A$, we conclude that $b - c = 0_A$ from $a \neq 0_A$ (there are no zero divisors). On the other side, if the cancellation law is valid, assume that $ab = 0_A$ for some $b \in A$. As $a0_A = 0_A$ we have that $ab = a0_A$ and then $b = 0_A$ (due to the cancellation law). ■

Definition For every natural number n and any element $a \in A$, we write

$$na := a + a + \cdots + a \ (n \text{ times})$$

(notice that n is a natural number, so this is not necessarily the operation of A, which is an abstract ring). We define the *characteristic* of a ring A as the least natural n such that $na = 0$ for every $a \in A$. If there is no n with this property, we define the characteristic as 0.

Lemma Let A be a ring with identity. If 1_A has order n (as an element of the group A under addition), then the characteristic of A is n.

Proof If 1_A is of order n then n is the least positive integer such that $n1 = 0$. Then, for all $a \in A$,

$$na = n(1_A a) = 1_A a + \cdots + 1_A a = (\underbrace{1_A + \cdots + 1_A}_{n \text{ times}})a = 0_A a = 0_A\,.$$

On the other side, if there is no positive n such that $n1_A = 0_A$, the ring characteristic is zero (because the property must be valid for all $a \in A$, and the property must hold for $a = 1_A$) ■

Proposition The characteristic of an integral domain is either a prime number or zero.

Proof Let A be an integral ring and let us assume that the characteristic of A is n with $n \neq 0$. If n is not prime, then $n = qd$ where $1 < q < n$ and $1 < d < n$. Due to the previous lemma, we must only consider the case $n1_A = 0_A$. We have that $0_A = n1_A = (qd)1_A = (q1_A)(d1_A)$. As there are no zero divisors in A, we conclude that either $q1_A = 0_A$ or $d1_A = 0_A$. In that case the characteristic of A would be strictly less than n, a contradiction. We conclude that n has no non-trivial divisors, and then it must be prime.■

7.4. Unique Factorization Domains

Definition (ASSOCIATE ELEMENTS) If A is a conmutative ring with identity element and a, b belong to A, a and b are *associates* if there exists an invertible element u such that $a = ub$.

Definition (PRIMES AND IRREDUCIBLE ELEMENTS) Let D be an integral domain. An element $q \in D$, $q \neq 0$, is called *irreducible* if for any factorization $q = ab$, we have that some of the factors (a or b) is invertible. An element $p \in D$ is *prime* if every time that $p|ab$ holds, then we have that $p|a$ or $p|b$.

The meaning of irreducible element and the one of prime are not equivalent in a general context, even though they are the same for integers. For polynomials, the most useful definition is the one of irreducible element.

The Fundamental Theorem of Arithmetic establishes that every positive integer $n > 1$ can be *factorized* as a product of prime numbers $p_1 \cdots p_k$, where p_i's can be repeated. These factorizations are unique up to reordering of the factors.

Definition (UNIQUE FACTORIZATION DOMAIN) We say that an integral domain D is a *unique factorization domain* UFD if D satisfies the following properties:
1. Let $a \in D$, $a \neq 0_D$ and not invertible. Then a can be written as a product of irreducible elements in D.
2. Let $a = p_1 \cdots p_r = q_1 \cdots q_s$, where the p_i's and q_i's are irreducible. Then $r = s$ and there exists an index permutation π such that p_i and $q_{\pi(j)}$ are associate , $j = 1, \ldots, r$.

8. Polynomials

Bibliography: [2, 3, 4].

Polynomials with coefficients in the real numbers (and functions in general) have a ring structure with usual sum and multiplication:

$$\begin{aligned} (p+q)(x) &= p(x)+q(x)\,, \\ (pq)(x) &= p(x)q(x)\,. \end{aligned}$$

For a generic x, the expressions on the right hand side return polynomials if p and q are also polynomials. The same happens with other "operators" as derivation and integration.

The basic operations over polynomials can be performed when the coefficients belong to a conmutative ring with identity element, but in order to have a well defined *division algorithm*, we need to divide coefficients, and it is then natural to ask them to belong to a *field*.

We will study abstract properties of polynomials, in the sense that we are not particularly interested in *evaluation* of the polynomial as a function of a number x, but more in its properties as a ring: factoring, division, Euclid's algorithm, etc. A very useful guide to not get lost in this subject is the following table that can be found in [3] (but sligthlty modified here). It shows a clear parallelism between $\mathbb{Z}$ and the ring of polynomials $K[x]$ with coefficients in a field K:

$a, b \in \mathbb{Z}$	$f, g \in K[x]$
absolute value of a	"degree" of the polynomial f
Invertibles: numbers ± 1	Invertibles: polynomials of 0 degree
Every integer can be written as: $a = a_0 + a_1 10 + \cdots + a_n 10^n$	Every polynomial can be written as: $f(x) = a_0 + a_1 x + \cdots + a_n x^n$
$b\|a$ if $a = bq$ for some integer q	$g\|f$ if $f = gq$ for some $q \in K[x]$
There exist q, $r \in \mathbb{Z}$ such that $a = bq + r$ with $0 \leq r < \|b\|$	There exist q, $r \in K[x]$ such that $f = gq + r$ with $\deg(r) < \deg(g)$
$\mathbb{Z}$ is an integral domain	$K[x]$ is an integral domain
a, b have a gcd d, that can be written as $d = ax + by$	f and g have a gcd $d \in K[x]$, s.t. $d = fq + gp$ with $q, p \in K[x]$
p is prime if it has no proper divisors	f is irreducible if it has no proper divisors
Every $z > 0$ is a "unique" product of primes	Every $f \in K[x]$, $\deg(f) > 0$ is a "unique" product of irreducibles

In what follows we assume that A is a conmutative ring with unity, and we can keep in mind that can be identified with $\mathbb{Z}$ or $\mathbb{Z}_n$ (if it is finite). K will be a field that can be identified with $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Z}_p$ for p prime when we consider a finite field.

8.1. Polynomials

An expression given by

$$p(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \cdots + a_n x^n$$

with $a_i \in A$, and $a_n \neq 0$ (the zero element of the conmutative ring A), is a *polynomial over the ring A and indeterminate x*. The "powers" of x have a formal meaning. To understand correctly the mathematical meaning of a polynomial, we must not identify the x as a variable, but as a formal element that may be replaced by some concrete value, when the powers and the products by elements of A are well defined. It is more convenient to think about x as an indeterminate element.

The polynomial is formally defined by the list of coefficients a_i and those coefficients determine its properties inside the ring of polynomials. The coefficient a_n is the *leading coefficient*. A polynomial is *monic* when its leading coefficient is 1 (the multiplicative unity of the ring A). The polynomial always has a *finite* amount of terms (a series of infinite terms is not a polynomial), so, the leading coefficient corresponds to the greatest n such that $a_n \neq 0$; that n is the *degree of the polynomial*. If the leading coefficient is a_0, then $a_0 \neq 0$ and $a_i = 0$ for all $i > 0$, the degree of the polynomial is zero. If all the coefficients are zero (p is identically zero), the degree of the polynomial is defined as $-\infty$.

Notation The set of polynomials over the ring A is written as $A[x]$. The degree of a polynomial p is denoted $\deg(p)$.

Two polynomials are considered equal when they have identical coefficients; that is, if

$$\begin{aligned} p(x) &= \sum_{i=0}^{m} a_i x^i \\ q(x) &= \sum_{i=0}^{n} b_i x^i \end{aligned}$$

then $p = q$ if and only if $a_i = b_i$ for every $i \geq 0$. Notice that the set of polynomials contains polynomials of *arbitrarily large degree*; that is, even though each polynomial has a finite degree, the set of polynomials as a whole has terms with arbitrarily large powers (similarly, the integers have a finite decimal representation, but we can find numbers with as large as we want). The operations sum and multiplication of polynomials are defined as:

Definition If p, q are polynomials in $A[x]$ the *sum* $p+q$ is a polynomial with coefficients $c_i = a_i + b_i$ for every $i \geq 0$.
The product pq has coefficients $c_i = \sum_{k=0}^{i} a_k b_{i-k}$, for $i = 0, \ldots, m+n$ that is obtained by grouping the powers i of x in the product $p(x)q(x)$. Notice that the degree of pq cannot exceed the sum of the degree of p and q, and some of the coefficients may be zero.

Theorem Let A be a conmutative ring with identity. Then $A[x]$ is a conmutative ring with identity.

Proof Exercise.

Proposition Let p and q be polynomials in $A[x]$, where A is an integral domain. Then $\deg(pq) = m+n$, where $\deg(p) = m$, $\deg(q) = n$. Moreover, $A[x]$ is an integral domain.

Proof If $p \neq 0$ and $q \neq 0$, then $pq \neq 0$ because $c_{m+n} = a_m b_n \neq 0$, then, if $pq = 0$ one of them must be null. We conclude that $A[x]$ is an integral domain■

Theorem Let A be a conmutative ring with identity and $\alpha \in A$. The map $\phi(\alpha) : A[x] \to A$ defined by

$$\phi_\alpha(p) = p(\alpha) = a_n\alpha^n + \cdots + a_1\alpha + a_0,$$

is a ring homomorphism, where $p(x) = a_n x^n + \cdots + a_1 x + a_0$.

Proof Exercise.

Definition The mapping $\phi_\alpha : A[x] \to A$ of the previous theorem is called *evaluation homomorphism in* α.

8.2. The division algorithm

For integers, the division algorithm says that if a and b belong to $\mathbb{Z}$, such that $b > 0$, then two unique integers q and r exist such that $a = bq + r$ and $0 \leq r < b$. For polynomials we have a similar property with several important consequences.

Theorem (Division Algorithm for polynomials) Let f and g be polynomials in $K[x]$[1], where $g \not\equiv 0$ (that is, is not the null polynomial). Then, exists polynomials q and r (quotient and remainder) in $K[x]$ such that $f = gq + r$, where we have that $\deg(r) < \deg(g)$. The polynomials q and r are unique with respect to these properties.

Proof If f were null, $q = 0$ and $r = 0$ is a solution (it has degree strictly lower than g, because $g \not\equiv 0$). If the degree of f is n and the degree of g is m, for $n < m$ we can write $q = 0$, $r = f$. We can see now the case $n \geq m$ by induction over m. Let

$$\begin{aligned} f(x) &= a_n x^n + \cdots + a_0 \\ g(x) &= b_m x^m + \cdots + b_0 \end{aligned}$$

[1] In this chapter, K is a field that we may identify with the rationals, the reals, the complex numbers or the integers modulo p for p prime.

If we divide g by b_m, and we multiply by the monomial $a_n x^{n-m}$ we define another polynomial with degree strictly lower than $\deg(f)$:

$$\tilde{f}(x) = f(x) - \frac{a_n}{b_m} x^{n-m} g(x)$$

Clearly we have that $\deg(\tilde{f}) < n$. This means that we may apply the inductive hypothesis to $\tilde{f}$:

$$\tilde{f} = g\tilde{q} + r\,,$$

where r has degree strictly lower than $\deg(g)$. This allows to write f in the following way:

$$f(x) = \tilde{f}(x) + \frac{a_n}{b_m} x^{n-m} g(x) = \left(\tilde{q}(x) + \frac{a_n}{b_m} x^{n-m}\right) g(x) + r(x)\,,$$

that is, by defining $q(x) := \tilde{q}(x) + \frac{a_n}{b_m} x^{n-m}$ the problem is solved. Let us show now uniqueness. Assume that there are two pairs: q_1, r_1 and q_2, r_2 that satisfy the requirements of the theorem, that is:

$$\begin{aligned} f &= gq_1 + r_1 \\ f &= gq_2 + r_2 \end{aligned}$$

with r_i of degree strictly less than $\deg(g)$ for $i = 1, 2$. Notice that we cannot have $q_1 = q_2$ and $r_1 \neq r_2$, that is, $(q_1, r_1) \neq (q_2, r_2) \Rightarrow q_1 \neq q_2$ and $r_1 \neq r_2$. By calculating the difference between them, we find:

$$0 = g(q_1 - q_2) + (r_1 - r_2) \Rightarrow g(q_1 - q_2) = r_2 - r_1\,.$$

If $q_1 \neq q_2$ the degree of $g(q_1 - q_2)$ is at least equal to the degree of g (if $q_1 - q_2$ has degree zero). On the other hand, the degrees of r_1 and r_2 are strictly smaller than the degree of g, so their difference has a degree strictly less than the one of g. The only way to satisfy the identity is with $q_1 - q_2 = 0$ and $r_1 - r_2 = 0$ ■

Definition Let p be a polynomial in $K[x]$ and $\alpha \in K$. We say that α *is a zero, or a root* of p if $p(\alpha) = 0$ (that is, $p(\alpha)$ is equal to the additive neutral element in K).

Theorem (REMAINDER) Let $p \in K[x]$ be a non-zero polynomial and let $c \in F$. Then, there exists a polynomial $q \in F[x]$ such that

$$p(x) = q(x)\,(x - c) + p(c)\,.$$

Proof We divide p by $(x - c)$, finding a remainder r such that $\deg(r) < \deg(x - c) = 1$, then $r \in K$. By evaluating the polynomial in $x = c$ we have that

$$p(c) = (c - c)q(c) + r \Rightarrow p(c) = r\,.$$

■

Corollary Let $p \in K[x]$ with $p \not\equiv 0$. α is a root of p if and only if $x - \alpha$ is a factor of p.

Proof By applying the remainder theorem, we have that, as α is a root of p, the remainder $r = p(\alpha)$ after division of p by $(x - \alpha)$ is zero, and then:

$$p(x) = (x - \alpha)q(x)\,.$$

On the other hand, if $x - \alpha$ is a factor of p, is is evident that α is a root of p ■

Corollary Let K be a field and $p \in K[x]$ with $p \not\equiv 0$, and $n = \deg(p)$. Then, p can have at most n different zeros in K.

Proof By induction on the degree of p. If $n = 0$, as $p \not\equiv 0$ we have that p is a non-zero constant, so it has no roots in K. Assume that the statement is true for polynomials of degree less or equal $n-1$. If α is a root of p, we consider the factorization:

$$p(x) = (x - \alpha)q(x)\,.$$

As q has degree $n-1$, it has at most $n-1$ roots, so p has at most n roots (the roots of q and α) ■

8.3. Greatest common divisor, Bézout identity and Euclid's algorithm

Definition Consider p, $q \in K[x]$. A monic polynomial $d \in K[x]$ is a *greatest common divisor* of p and q if d divides exactly p and q and, further, if any other polynomial $\tilde{d}$ has the same property, then $\tilde{d}|d$. We write $d = \gcd(p, q)$. Two polynomials p and q are *coprime* or *relative primes* if $\gcd(p, q) = 1$ (the monic polynomial of degree zero).

Theorem (Bézout identity for polynomials) Assume that d is the maximum common divisor of p and q in $K[x]$[2]. Then, there exist polynomials a and b such that

$$d(x) = a(x)p(x) + b(x)q(x).$$

Moreover, the gcd of two polynomials is unique.

Proof We select d as the monic polynomial of *least degree* from the set:

$$S = \{c\,(x) :\ c\,(x) = f(x)p(x) + g(x)q(x),\ c \not\equiv 0\,,\ f, g \in K[x]\}\,.$$

This can be done due to the fact that the set of *degrees* of monic polynomials are positive integers, and they must have a least element. This polynomial d then can be written as $d(x) = a(x)p(x) + b(x)q(x)$ for some polynomials a, b. To show that d divides p, we consider the division

$$p = hd + r \Rightarrow r = p - hd$$

[2]Remember that K is a field and we identify it with the rationals, the reals, the complex or the integers modulo p for p prime.

were h, r are the quotient and remainder obtained from the division algorithm. Using the expression for d we have that

$$r = p - h\,(ap + bq) = (1 - ha)\,p - hbq\,.$$

But then r is also a combination of p and q with degree strictly less than d. The only possilbe way is that $r \equiv 0$ because we assumed that d has the least degree among the non-zero combinations. So we have that $r \equiv 0$ and d divides p. The case of q is dealed in the same manner.
Let us prove now that it must be the *greatest* common divisor. If $\tilde{d}|p$ and $\tilde{d}|q$, we can write, for some polynomials f and g: $p = f\tilde{d}$ and $q = g\tilde{d}$, then:

$$d = af\tilde{d} + bg\tilde{d} = (af + bg)\,\tilde{d}$$

then $\tilde{d}|d$.
We see now that d is *unique*. If d' divides p and q, and for any other polinomial h with the same property, it must hold that $h|d'$, then as d divides p and q we must have that $d|d'$. Previously, we showed that d' must divide d. Then:

$$d = qd' \Rightarrow \deg(d) = \deg(q) + \deg(d') \Rightarrow \deg(d) \geq \deg(d')$$

and we can obtain the opposite inequality by the same argument. This implies that they have the same degree and that the degree of q must be zero. Then, h is a constant. As both polynomials are monic, then $h \equiv 1$ and $d \equiv d'$ ■

Proposition Given p, f, $g \in K[x]$, if $\gcd(p, f) = 1$ (that is, if p and f are coprime) and $p|fg$ then $p|g$.

Proof We saw the same property for integers. As p and f are coprime, by Bézout identity we can find polynomials a and b such that:

$$1 = ap + bf\,.$$

Then

$$g = apg + bfg$$

As $p|fg$ there must be a q such that $fg = qp$. Then:

$$g = apg + bqp = (ag + bq)\,p$$

so, we have that $p|g$ ■

Algorithm (Euclid's algorithm for polynomials) We already know the procedure from the one used for integers. There is a complete analogy with the polynomial case, and we only nned to take into account that the gcd must be a monic polynomial. If we want to calculate the gcd of f and g, (assumig that the degree of f is greater or equal to the one of g) we apply the division algorithm and find

$$f = qg + r$$

If d divides f and g, then it divides r, so we have that $\gcd(f,g) = \gcd(g,r)$ and in this way we reduce the degree of the polynomials involved, until we have $r \equiv 0$. The remainders can be written in a table:
If we divide f by g, we withe $f = q_1 g + r_1$ (the 1 indicates the row of the table), so we have that $r_1 = f - q_1 g$, then the first line is:
(i) $\boxed{r_1 \mid 1 \mid -q_1}$ where 1 refers to the coefficient of f and $-q_1$ for the one of g.
Then we divide g by r_1 and obtain: $g = q_2 r_1 + r_2 \Rightarrow r_2 = g - q_2 r_1$. We use the former coefficients to fill the new row:
(ii) $\boxed{r_2 \mid -q_2 \mid 1 + q_1 q_2}$
Following this procedure, in each step we obtain the coefficients of f and g and find the quotients. In a given step k we have:
(k) $\boxed{r_k \mid \alpha_k \mid \beta_k}$ with $r_{k-2} = r_{k-1} q_k + r_k$, that is $r_k = r_{k-2} - r_{k-1} q_k$. This gives $(\alpha_{k-2} - \alpha_{k-1} q_k)\, f + (\beta_{k-2} - \beta_{k-1} q_k)\, g$, and then $\alpha_k = \alpha_{k-2} - \alpha_{k-1} q_k$ and $\beta_k = \beta_{k-2} - \beta_{k-1} q_k$, until the step n in which:
(n) $\boxed{r_n \mid \alpha_n \mid \beta_n}$ and
(n+1) $\boxed{0 \mid \alpha_{n+1} \mid \beta_{n+1}}$
Then the gcd is r_n and we have a Bézout identity:

$$d = \alpha_n f + \beta_n g\,.$$

Remark Euclid's algorithm is a *constructive* proof of Bézout identity.

8.4. Irreducible polynomials and factorization

Definition A polynomial with degree greater than zero (ie. non-constant) $f \in K[x]$ is *irreducible* on the field K if f cannot be factorised as a product of two polynomials g and h in $K[x]$, where the degrees of g and h are both smaller than the degree of f. Irreducible polynomials in polynomial rings play the same role as the prime numbers within the integers.

Examples A polynomial of degree 1 is irreducible, because the product of two polynomials of zero degree must be a polynomial o zero degree (the factors must have a smaller degree). If we take for example $a\,(x-1)$, with $a \neq 0$, this does not mean that $(ax - a)$ is reducible, because the degree of $x - 1$ is not smaller than the grade of the factorised polynomial. This is compatible with the abstract definition of being irreducible, because $ax - a = a(x-1)$, where a is invertible (because it is a non-zero element of a field).
The polynomial $x^2 - 2$ in $\mathbb{Q}[x]$ is irreducible, because it cannot be factorised over the rationals. If it were so, it must have two factors of degree 1, and then two roots in $\mathbb{Q}$. In a similar manner, $x^2 + 1$ is irreducible over the real field of numbers.
On the field $\mathbb{Z}_2$, the polynomial $x^2 + x + 1$ is irreducible, but in $\mathbb{Z}_3$ it has a root (1) and can be factorised as $(x-1)^2 = x^2 - 2x + 1 = x^2 + x + 1$ because $-2 \equiv_3 1$.

Lemma A non constant polynomial $p \in K[x]$ is irreducible on K if and only if for any two polynomials f, $g \in K[x]$, if $p|fg$ then $p|f$ or $p|g$.

Proof Let us show first that p irreducible $\Rightarrow p|fg$, implies $p|f$ or $p|g$. Assume that $p|fg$ but does not divide f. Being p irreducible, the gcd of p and f is 1 (because p does not divide f), then p and f are relatively prime. Using a previous proposition, we conclude that $p|g$. Then, it must divide one of them.
Let us check $\Leftarrow$. We assume that every time p divides a product fg, p must divide f or g, so we must check that p is irreducible. We will show that if p is not irreducible this cannot hold, because if $p = uv$ with u, $v \in K[x]$, both with degree strictly lower than $\deg(p)$, then p divides uv, but p cannot divide u nor v ■

Proposition A polynomial of degree 2 or 3 is irreducible on the field K if and only if it has no roots in K.

Proof The only way to factorise the polynomial with factors strictly lower in degree is with two of degree 1 (if the original polynomial is of degree 2) or one of degree 1 and other of degree 2 (if the original polynomial has degree 3). In both cases, the polynomial has a factor of degree 1, so it must have a root in K■

Note In the general case it is not enough to check if the polynomial has roots in the given field. For example, in $\mathbb{Q}[x]$ the polynomial $(x^4 - 4) = (x^2 - 2)(x^2 + 2)$ has no roots in $\mathbb{Q}$, but is reducible.

While studying roots and polynomial factorisation it is useful to know if there are repeated roots or factors. The *derivative* p' of the polynomial p can be used to find them. It is possible to formally define the derivative of a polynomial without reference to the limit process of Calculus.

Definition If $p(x) = a_0 + a_1x + \cdots + a_nx^n$ is a polynomial of degree n in $K[x]$, the *derivative* p' is defined as:

$$p'(x) = a_1 + 2a_2x + \cdots + nx^{n-1}\,.$$

Notice that if p has degree zero, its derivative is the null polynomial, with degree $-\infty$.

Lemma The rule for product derivative is valid in $K[x]$:

$$(fg)' = f'g + fg'\,.$$

Proof By induction. If f has degree 0 and g has arbitrary degree m, $f = a_0$ and $f' = 0$. Let $g = b_0 + \cdots + b_mx^m$, so $fg = a_0g = a_0b_0 + a_0b_1x + \cdots + a_0x^m$, clearly

$$(fg)' = a_0b_1 + a_02b_2x + \cdots + a_0mb_mx^{m-1} = a_0g' = fg' + f'g\,.$$

Given n, we assume that the formula is valid for every f of degree less than n, and we will see that it is valid for degree n. Let $f = a_0 + \cdots + a_nx^n$. We

write $fg = (a_n x^n g) + g\tilde{f}$, where $\tilde{f} = f - a_n x^n$. $a_n x^n g = a_n b_0 x^n + a_n b_1 x^{n+1} + \cdots + a_n b_m x^{n+m}$. Then $(a_n x^n g)' = b_0 n a_n x^{n-1} + \cdots + b_m (n+m) a_n x^{n+m-1} = (a_n x^n)' g + (a_n x^n) g'$. As $\tilde{f}$ has degree strictly lower than n, we have that $\left(\tilde{f}g\right)' = \tilde{f}'g + \tilde{f}g'$. Gathering all the terms we obtain:

$$\begin{aligned}(fg)' &= \left(\tilde{f}g + a_n x^n g\right)' = \tilde{f}'g + \tilde{f}g' + n a_n x^{n-1} g + a_n x^n g' \\ &= \left(\tilde{f}' + n a_n x^n\right) g + \left(\tilde{f} + a_n x^n\right) g' \\ &= f'g + fg' .\end{aligned}$$

■

Definition Let $p \in K[x]$. An element $\alpha \in K$ is a *root of multiplicity* $m \geq 1$ of p if

$$(x-\alpha)^m \,|\, p \quad \text{but} \quad (x-\alpha)^{m+1} \not| p .$$

Proposition A non-constant polynomial p on K has no repeated factors if and only if $\gcd(p, p') = 1$, that is, if p and p' are coprime.

Proof It is equivalent to show that p has repeated factors if and only if p and p' are not coprime. Assume that the greatest common divisior of p and p' is d sudch that $\deg(d) > 1$. If f is an irreducible factor of d (that can be d itself), we have that $f|p$ and $f|p'$. That is $p = af$, $p' = bf$, for some a, b polynomials. On the other hand, due to the rule for product derivative $p' = a'f + af' = bf$. This shows that f divides af' because $af' = (b - a')f$. Being f irreducible, it divides the product af' and f does not divide f' (because f' has lower degree), then f must divide a, say $a = qf$, and then $p = af = qf^2$. We conclude that p has a repeated factor.
On the other hand, if p has a repeated factor (of degree greater than 1), then $p = f^n q$ with $n > 1$. And we have that

$$p' = (f^n)' q + f^n q' = n f^{n-1} f' q + f^n q' = f^{n-1} \left(n f' q + f q'\right)$$

Having $n > 1$, p and p' have a common divisor f, and then they are not coprime
■

The following theorem can be proved in a very similar way to the integer case, replacing "irreducible polynomials" by "prime numbers, and "degree" by "abslulte value". The theorem states that the set of polynomials with coefficients in a field are a *unique factorization domain.*

Theorem (UNIQUE FACTORIZATON) Every non constant polynomial with coefficients in a field K can be factorised as a product of an element in K and a product of monic, irreducible polynomials over the field K. This factorisation is unique, except for factor reordering.

8.5. Ideals and Congruences in $K[x]$

Theorem Let I be a subset of $K[x]$ satisfying the following conditions:
(i) I has at least a nonzero polynomial;
(ii) If f , $g \in I$, then $f + g \in I$;
(iii) If $f \in I$ and $q \in K[x]$, then $qf \in I$.
Then, if d is a non zero polynomial in I, of least degree, I can be characterised as:

$$I = \{p \in K[x] : p = qd, \text{ for some } q \in K[x]\}$$

(that is, every element of I has d as a factor).

Proof If I has a nonzero polynomial, then the set of positive integers given by the degrees of the polynomials in I has a least element m and there exists $d \in K[x]$ such that $\deg(d) = m$.
Every multiple qd must be in I, due to condition **(iii)**. We must check that every polynomial in I has this form, that is, if $f \in I$, then $f = qd$ for some $q \in K[x]$. We then apply the division algorithm to f and d:

$$f = qd + r$$

with $\deg(r) < \deg(d)$. As $d \in I$, then (due to **(iii)**) $qd \in I$ and so $-qd \in I$ (-1 is a polynomial in $K[x]$) then $f - qd \in I$. This means that r belongs to I and has a degree strictly lower than m. r must be the zero polynomial and then f is divisible by d ■

Definition p_1, p_2,..., $p_n \in I$ are generators of the ideal if

$$I = \{q_1p_1 + \cdots + q_np_n : q_1 \ldots q_n \in \mathbb{K}[x]\} .$$

Definition A *principal ideal* in $K[x]$ is an ideal "generated" by a polynomial $p \in K[x]$, that is:

$$\langle p \rangle = \{pq : q \in K[x]\} .$$

This set is similar to a subgroup in an abelian group. It can be verified that this set is a *subring* in $K[x]$. It is also similar to the subgroups of $\mathbb{Z}$, that are multiples of a fixed number: $\langle m \rangle = m\mathbb{Z} = \{mq : z \in \mathbb{Z}\}$. For the integers, we defined the group quotient $\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}_m$. That is, the classes defined by the subgroup $m\mathbb{Z}$ are the same as the classes "modulo" m, that have a sum and product operation. On the other hand, when m is a prime, the quotient *is a field*, because every non-zero element of $\mathbb{Z}_p$ when p is prime *has multiplicative inverse*.

For polynomials all these facts are almost identical. We only need to find the "quotient classes", their operation and (by means of Euclid's algorithm), their prime elements and the "quotient fields".

Definition Let p be a fixed polynomial in $K[x]$. If a, $b \in K[x]$, we say that they are *congruent modulo* p, denoted by $a \equiv b \pmod{p}$ if p divides $a - b$.
The set $\{b \in K[x] | b \equiv a \pmod{p}\}$ is the *congruence class* of $a \in K[x]$, and is

deoted by $[a]_p$ or $[a(x)]_{p(x)}$ (if we know that a, p are polynomials we may omit the x).

The set of all congruence classes modulo p is denoted by the quotient $K[x]/\langle p\rangle$.

Remark The quotient notation is due to the fact that we are working with the equivalence classes defined by $a \sim b \Leftrightarrow a - b \in \langle p\rangle$. Remember that $\langle p\rangle$ are all the "multiples" of p, that is, the polynomials qp with $q \in K[x]$. This is the same as claiming that $a - b$ is divisible by p.

While working with the integers modulo n, that is $\mathbb{Z}_n$, we considered a representative element of the class $[z]_n$ that satisfied $0 \leq z < n$. In a similar way, we can take the polynomial of least order among the class modulo $p(x)$ of a polynomial.

Proposition Let $a, p \in K[x]$ and $p \neq 0$. The congruence class $[a]_p$ contains exactly one polynomial r such that $\deg(r) < \deg(p)$.

Proof Applying the division algorithm:

$$a = qp + r$$

We have that $\deg(r) < \deg(p)$ and

$$a - r = qp$$

that is, p divides $a - r$, r is a representative element of the a class, with a degree strictly lower than p. According to the division algorithm, r is the unique polynomial with this property. Notice that, if $r = 0$ the degree of r is $-\infty$ and it has (trivially) lower degree than p. In this case, if a is divisible by p, belongs to $[0]_p$, that is, the class of the zero polynomial ■

Proposition Let $p \neq 0$ be a polynomial in $K[x]$. Given a, b, c, $d \in K[x]$, the following statements are true:

(a) If $a \equiv_p c$ and $b \equiv_p d$, then $a + b \equiv_p c + d$ and $ab \equiv_p cd$.

(b) If $ab \equiv_p ac$ and $\gcd(a, p) = 1$, then

$$b \equiv_p c\,.$$

Proof Is very similar to the proof for integers. We prove **(b)**. If ab and ac are congruent modulo p, we have that

$$ab - ac = qp \Rightarrow a\,(b - c) = qp\,.$$

As a and p are coprime, we have seen that then p must divide $b - c$ (by using Bézout's identity), then $b \equiv_p c$ ■

This propositon allows us to define the sum and product operations *over equivalence classes of* $K[x]/\langle p\rangle$.

Proposition Consider $p \neq 0$ such that $p \in K[x]$. Given $a \in K[x]$, the class $[a]_p \in K[x]/\langle p\rangle$ has a multiplicative inverse if and only if $\gcd(a, p) = 1$.

Proof $[a]_p$ has a multiplicative inverse if there exists a class $[b]_p$ with $b \in K[x]$ such that $[a]_p[b]_p = [1]_p$. That is, $ab - 1$ must be divisible by p

$$ab - 1 = qp$$

then

$$ab - qp = 1\,,$$

and the polynomials b and q with this property exist if and only if the greatest common divisor of a and p is the constant polynomial 1. The polynomial b defines the class of the inverse of $[a]_p$, that is $[a]_p^{-1} = [b]_p$ and can be found by means of Euclid's algorithm ■

Theorem Let $p \in K[x]$ be a non constant polynomial. $K[x]/\langle p\rangle$ is a *field* if and only if p is irreducible on K.

Proof Sum and multiplication are well defined. Moreover, the operations are commutative, associative and distributive. The additive neutral element is $[0]_p$ and the multiplicative neutral is $[1]_p$. It only remains to verify that every element different from $[0]_p$ has a multiplicative inverse. To this end, consider the class of a nonzero polynomial, and inside the class, the representative element a with least degree (strictly lower than p). In order to find the inverse of $[a]_p$ we must show that the gcd of a and p is 1. Being p irreducible, there is no polynomial that divides it, and then $\gcd(a, p) = 1$. By Bézout's identity, there exist polynomials $h(x)$, $g(x)$ such that

$$h(x)\,a(x) + g(x)\,p(x) = 1 \Rightarrow 1 - h(x)\,a(x) = g(x)\,p(x)\ ,$$

that is, ha is congruent to 1 modulo p, then h is the inverse of a. This completes the proof that $K[x]$ is a field ■

To carry further the analogy between $K[x]$ and $\mathbb{Z}$, notice that $\langle p\rangle$ is a set closed under the sum (they are the multiples of p) and then they are an abelian group (with respect to sum). The quotient $K[x]/\langle p\rangle$ is also a group under the operation sum, in the same way as the quotient $\mathbb{Z}/n\mathbb{Z}$ (that we called $\mathbb{Z}_n$) is a group with respect to sum. As we already mentioned, $K[x]/\langle p\rangle$ has a multiplication too, and is a ring, but the nonzero elements, are not a multiplicative group (in the same way that the nonzero elements of $\mathbb{Z}_n$ do not form a group, unless n is prime), unless p is irreducible. The invertible elements $K[x]/\langle p\rangle$ are a multiplicative group, analogous to the group of units $\mathbb{Z}_n^*$.

Definition Given two fields K_1 and K_2, a mapping $\phi : K_1 \to K_2$ is a *field homomorphism* if

$$\phi(a + b) = \phi(a) + \phi(b)\,, \quad \text{and } \phi(ab) = \phi(a)\phi(b)\,, \quad \forall a, b \in F_1.$$

If ϕ is bijective, it is a *field isomorphism.*

The following proposition identifies K as a "subfield" of $K[x]/\langle p\rangle$.

Proposition If p is a non constant polynomial, the mapping $\phi : K \to K[x]/\langle p\rangle$, that sends each constant polynomial $a \in K$ to its corresponding class $[a]_p$ is an injective ring homomorphism, and then it is an isomorphism from K to its image $\{[a]_p : a \in K\}$. Moreover, if p is irreducible, it is a field homomorphism.

Proof We have that $\phi(a+b) = [a+b]_p = [a]_p+[b]_p = \phi(a)+\phi(b)$ and we have a similar identity with multiplication. We must check injectivity. This is almost evident, because the class is completely defined by the representative with degree lower than $\deg(p)$. That is, if a, b, $\in K$ and $[a]_p = [b]_p$, we have that $a-b$ is divisible by p. But p is not constant and $a-b$ has degree zero. To be divisible, the only possible case is $a-b=0$, that is the representative is the same element ■

We can think of $K[x]/\langle p\rangle$ as an *extension* of the field K. It is always possible to extend K in such a way that includes the roots of a given polynomial. This is Kronecker's theorem.

Teorema (KRONECKER) Let f be a non constant polynomial in $K[x]$. There is a field E, an extension of K, such that $\alpha \in E$ and $f(\alpha) = 0$.

Proof As f can be expanded as a product of irreducible factors, we consider one of these irreducible factors p. We look for an element α such that $p(\alpha) = 0$. As p is a factor of f, this will be enough for our purpose. We consider the field $E := K[x]/\langle p\rangle$. As we already know, the field K is isomorphic to a subfield of E, that is formed by the classes $[a]_p$ for $a \in K$. Let α be the congruence class $[x]_p$ (that is, the class of the monomial x with constant equal to zero). If $p(x) = a_0 + a_1x + \cdots + a_nx^n$, $a_i \in K$, we calculate $p(\alpha)$:

$$\begin{aligned} p(\alpha) &= a_0 + a_1[x]_p + \cdots + a_n[x]_p^n \\ &= [a_0 + a_1x + \cdots + a_nx^n]_p \\ &= [p]_p = [0]_p \end{aligned}$$

That is, E extends K as a field and contains a root of f ■

Corollary If f is a non constant polynomial in $K[x]$, there exist an extension of E of K where f can be factorised as a product of linear factors (that is, of degree 1).

Proof We factorise f using all the roots of f in K. Let f_1 be the remaining factor. We can extend K to a field E_1 in such a way that f_1 has a root in E_1. Then, we can write $f_1(x) = (x-u_1)f_2(x)$, and taking f_2 as an element of $E_1[x]$, we continue the procedure with f_2. We eventually build a field E that has enough roots to factorise f as a product of linear factors ■

Example (CONSTRUCTION OF THE COMPLEX NUMBERS) Take K as the field $\mathbb{R}$ of real numbers. The polynomial $p(x) := x^2+1$ is irreducible in $\mathbb{R}[x]$ and then $\mathbb{R}[x]/\langle x^2+1\rangle$ is a field with the operations defined between classes. Given that $[x^2+1]_p = [0]_p$, we have: $[x^2]_p = [x]_p^2 = [-1]_p$, that is, the class of x behaves as the complex number i, if we *identify the class of* -1 with the number -1 (by the proposition above). On the other hand, each class has a representative

element that is a polynomial of degree 1 with real coefficients, that is, each polynomial of the form $a + bx$ defines a different class. This, together with the fact that the class of x^2 is the same as the class of -1, shows that que quotient is a field "identical" (via isomorphism) to the field $\mathbb{C}$.

Example Consider the polynomial $f(x) = x^4 - x^2 - 2$ with coefficients in $K = \mathbb{Q}$. f can be factorised as $(x^2 - 2)(x^2 + 1)$, and as a first step we write $E_1 = \mathbb{Q}[x]/\langle x^2 - 2\rangle$, that is isomorphic to $\mathbb{Q}(\sqrt{2})$ (exercise). E_1 contains the roots $\pm\sqrt{2}$ of the factor $x^2 - 2$, but does not contain the roots $\pm i$ of the polynomial $x^2 + 1$, so that we consider a second extension $E_2 = E_1[x]/\langle x^2 + 1\rangle$. It can be proved that E_2 is isomorphic to the smaller extension of $\mathbb{Q}$ containing i and $\sqrt{2}$ denoted by $\mathbb{Q}(\sqrt{2}, i)$.

Example Consider $K = \mathbb{Z}_2$ and $p(x) = x^2 + x + 1$. p is irreducible in $\mathbb{Z}_2$ because it has no roots there and then we cannot factor it with linear factors. Then $\mathbb{Z}_2[x]/\langle x^2 + x + 1\rangle$ is a field. The congruence classes modulo p are represented by the polynomials 0, 1, x and $1+x$ (that is, all the polynomials of degree lower than 2 that we can construct with coefficients in $\mathbb{Z}_2$).

Property The quotient $K[x]/\langle p\rangle$, where $p(x)$ is an irreducible polynomial is a *vector space* of dimension n =degree of p, over the field K. To see this, take into account that every class is defined by the polynomials of degree $n - 1$ with coefficients in K:

$$\begin{aligned}[f]_p &= [a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}]_p \\ &= a_0[1]_p + a_1[x]_p + \cdots + a_{n-1}[x^{n-1}]_p\end{aligned}$$

That is, all the classes are generated with n coefficients that belong to K: $(a_0, \ldots, a_{n-1})$, so that the quotient has a vector space structure identical the one of K^n. In general, the dimension as a vector space of an extension E of a field K is called the *degree* of E over K and is denoted by $[E : K]$.

8.6. If K is a finite field

We are mainly interested in the case where K is some $\mathbb{Z}_p$ for p prime. If $q(x)$ is irreducible in $\mathbb{Z}_p[x]$, then $\mathbb{Z}_p[x]/\langle q(x)\rangle$ is a field and has p^n elements if n is the degree of the polynomial $q(x)$. It is possible to find irreducible polynomials of arbitrary degree $n > 0$. This guarantees the existence of fields with exactly p^n elements, with p prime being n an arbitrary positive integer. More generally, it is possible to prove that all the fields of order p^n are isomorphic, so a unique notation is used for them: $\mathbb{F}_{p^n}$, these are the so-called *Galois fields*. The multiplicative group of these fields (that is, the nonzero elements with multiplication) is denoted by $\mathbb{F}^*_{p^n}$.

Bibliography

[1] *Algebra for Computer Science*, Lars Gårding, Torbjörn Tambour. Universitext, Springer 1988.

[2] *Abstract Algebra: Theory and Applications,* Thomas Judson. Free download: http://abstract.ups.edu/download/aata-20140815.pdf.

[3] *Applied Abstract Algebra*, Rudolf Lidl & Günter Pilz. Springer 1998 (Second edition).

[4] *Abstract Algebra*, John Beachy & William D. Blair. Waveland Press 2006 (Third edition).