Tel Aviv University
The Blavatnik School of Computer Science
Winter 2018

Abstract Algebra in Theoretical Computer Science

Gil Cohen

Preface

Table of Contents

Lecture 2
The Fundamental Theorem of Algebra

In this lecture, we will discuss the Fundamental Theorem of Algebra while exploring the number systems $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. In retrospective, one can think of this process emerging so that the theorem will hold. The extension of some of these number systems from their prior will be abstracted later in the course and so it is beneficial to see the ideas involved on a familiar ground. I will mostly follow Chapters 1,2 of Stewart's excellent book on Galois Theory Stewart [2015].

2.1 The Familiar Number Systems

Solving polynomial equations, despite its boring reputation, has a fascinating history and required some significant psychological leaps from the very best of mathematicians (and it still does from the best of students). Slowly but surely, mathematicians extended their "number systems" when encountered with a problem expressed within the known number system whose solution was "outside" of it. In this section we briefly review this process. Some of the ideas that are required for extending these number systems will be abstracted later in the course and so it is beneficial to recall these ideas when applied at a familiar ground which we, at the very least, think we understand.

It all started with the set $\{1, 2, 3, \ldots\}$ which by itself is a completely non-trivial concept. It was highly abstract a few thousand years ago. It also didn't help that this set is infinite. Even today, many high school students are confused about the alleged paradox that every number is "finite" yet there are infinitely many of them. In this number system we can solve equations like $x + 1 = 2$. I will leave this as an exercise.

2.1.1 Enter zero

The acceptance of zero as a legitimate number took some getting used to. The ancient Greeks, for example, had no symbol for zero as they baffled with deep philosophical questions such as "how can nothing be something?". Zero was used as a placeholder quite early in positional number systems like we use today, but it was considered nothing more for years to come. We write $\mathbb{N} = \{0, 1, 2, \ldots\}$ for the *natural numbers*. Yes, we consider 0 to be a (very) natural number. In fact, when we come to formalize

the notion of a number system using an axiomatic approach, the existence of (the abstraction of) 0 will be one of the axioms. More so, it will be the *only* number we demand to exist within the number system.

2.1.2 Negative numbers

Don't get me started about the negatives which allows one to solve equations like $x + 1 = 0$. It suffices to say that even at 1759, the English mathematician Maseres wrote that negative numbers "darken the very whole doctorines of equations and make dark the things which are in their nature excessively obvious and simple". Leibniz is considered to be the first to systematically employ negative numbers. He did so for his development of Calculus. I don't know about you, but I always imagined that Calculus is light years away from any discussion about negative numbers. Anyhow, denote the *whole numbers* by $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. The letter $\mathbb{Z}$ comes from the German word "Zahl" which translates to a teller in English.

2.1.3 Be rational

What about $2x = 1$? Positive fractions seem to have been recognized earlier than zero and the negatives. However, there is some complexity involved in their formal definition. We are used to think of rational numbers as, well, numbers or more precisely as a pair of whole numbers. In particular, we write $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z},\ b \neq 0\}$. However, we identify some of the numbers in this set such as $\frac{1}{2}$ and $\frac{2}{4}$. So, in fact, a rational number is not quite a pair of $\mathbb{Z}$ elements but rather a set of such pairs. More precisely, a rational number is an equivalent class with respect to some equivalent relation. However, we are so used to this that we suppress this fact and, in particular, write things like $\mathbb{Z} \subset \mathbb{Q}$ which formally does not make much sense. What we actually mean is that there is a copy of $\mathbb{Z}$ "embedded" in $\mathbb{Q}$. This copy is given by $\{\frac{a}{1} \mid a \in \mathbb{Z}\}$ and it behaves like $\mathbb{Z}$ when we add and multiply unlike, say, $\{\frac{1}{a} \mid a \in \mathbb{Z}\} \cup \{0\}$.

Exercise 1. The Egyptians only considered fractions of the form $\frac{1}{a}$ for $a \in \{1, 2, 3, \ldots\}$ (and $\frac{2}{3}$ but let's ignore that one). One nice and not completely trivial fact is that any fraction $\frac{a}{b}$ with $1 \leq a \leq b$ can be written as a finite sum of distinct Egyptian fractions. Can you prove that?

Later in the course we will abstract this process of taking a number system like $\mathbb{Z}$, some of whose elements cannot be inverted, and "embed" it in a bigger number system that is closed to inversion.

2.1.4 Get real

Attempts made by ancient mathematicians who recognized only $\mathbb{Q}$ as the set of numbers to solve $x^2 = 2$ is a famous story in the history of Mathematics. Once, again, the realization that this is impossible came as a philosophical shock.

Exercise 2. Here is a lesser-known proof sketch for the insolvability of $x^2 = 2$ in $\mathbb{Q}$. Try to fill in the details. Assume by way of contradiction that $\frac{a}{b}$ is a solution to $x^2 = 2$ with $a, b \in \mathbb{N}$ and b minimal among all such solutions. Consider now the expression $\frac{2b-a}{a-b}$.

Extending $\mathbb{Q}$ to $\mathbb{R}$ is completely non-trivial. It involves taking the topological closure of $\mathbb{Q}$ with respect to the natural metric and by that close the (many many) "wholes" in $\mathbb{Q}$. This is more or less done by adjoining the limits of all convergent sequences in $\mathbb{Q}$. Anyhow, whatever $\mathbb{R}$ is, it is fairly safe to say that we all feel comfortable with it. We don't call them real numbers for nothing!

2.1.5 Complex numbers

What about solving $x^2 + 1 = 0$? We are all programmed to shout i (or $\pm i$) but deep inside one might have the feeling that i is just a made up symbol–a cheat if you will. I mean, $\sqrt{2}$ I can get–it is the limit of a sequence of approximate solutions to $x^2 = 2$. But i is just, well, not real... Like the zero and the negative numbers, i wasn't greeted with a smile by humankind. It was more like, well, we really need this guy to solve equations, but it was considered as this formal symbol that one can manipulate but dare not consider as "real".

Let's elaborate on that. We all know how to solve the general quadratic equation $ax^2 + bx + c = 0$. We have this neatly wrap expression for the solutions

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

which I'm positive you can cook up by yourself. This formula, expressed quite differently, was known already to the Babylonians some 3600 years ago. Applying this to $x^2 + 1 = 0$ doesn't give any meaningful answer in $\mathbb{R}$ as the $\sqrt{\cdot}$ is applied to a negative number. This wasn't a problem to i-non-believers. For them, it was simply Math's way of telling us that there is no solution.

i came to hunt the human race when people were finally able to solve cubic equations. It turns out that there is a general solution to such equations and one can derive it in a page or two (see Stewart's book). However, there is a significant amount of trickery involved and it was an open problem to come up with a solution for quite some time. It was only at around 1535 that the general cubic equation was solved by

Fontana (nicknamed Tartaglia). First, using some standard trickery, one can reduce the general cubic equation to the form $x^3 + px + q = 0$. A general solution is then given by, get ready for this,

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Impressive no doubt. But, here is the catch. If we apply this to $x^3 - 15x - 4 = 0$ which clearly has a solution $x = 4$, we get $x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$. Where is our beloved 4? Turns out that if you are willing to consider i as if it was a legitimate number, assuming all rules of arithmetics apply to him, you can extract 4 out of this mess.

You see, it is not just that $x^3 - 15x - 4$ has solutions outside of the reals which you may or may not choose to consider as real. It is that even if these solutions are real as 4, our way of finding them gets out of $\mathbb{R}$ before landing back safely. You might not be so impressed. After all, this is just one way of finding a solution. Perhaps the undesired visit of i is due to the algorithm (the formula) not the problem itself. Well, turns outs that one can prove, in some formal sense, that any solution that is expressed by radicals (square roots, cubic roots, etc) will go through i even in some cases in which all roots are real. Indeed, Mathematics is trying to tell us something... Soon enough we'll start talking about "field extensions". The uncomfortable feeling we may have had with i–adding this artificial solution–will come to hunt us again. So, we better surface these feelings at a familiar ground.

At any rate, we define $\mathbb{C} = \{a + ib \mid a, b \in R\}$ where addition and multiplication are given by "extending" these operations from $\mathbb{R}$ together with the rule $i^2 = -1$. So, multiplication is given by

$$\begin{aligned}(a + bi)(c + di) &= ac + adi + bci + bdi^2 \\ &= (ac - bd) + (ad + bc)i.\end{aligned}$$

Going back to our friend, $\sqrt{2}$. Come to think of it, if something is not real then it is $\sqrt{2}$. I mean it is an endless pattern-less string of digits. There is not enough atoms in the universe to represent this idealized number. So, I claim you have never seen the *real* $\sqrt{2}$ in your life!

2.2 Will this ever end?

One of the many cool features of $\mathbb{C}$ is that it is the end of this game. $\mathbb{C}$ has the remarkable property that *any* polynomial equation with coefficients in $\mathbb{C}$ has *all* of its solutions in $\mathbb{C}$. That's a great deal! We added only this single weird symbol i so as to obtain/invent/discover, you choose, a solution to the specific simple equation $x^2 + 1 = 0$ and what I'm saying is that by doing that, we got all solutions to all polynomial equations for free even if the coefficients have i's in them! Later in the course we will refer to number systems that have this property *algebraically closed*. When I say that $\mathbb{C}$ is the end of the game, I don't mean that $\mathbb{C}$ is the only number system with this property. I mean that it is the only one if you start from $\mathbb{R}$.
This property of $\mathbb{C}$ is given by The Fundamental Theorem of Algebra. To state it, recall that a solution to a polynomial equation $p(x) = 0$ is called a *root* of p.

Theorem 2.3 (The Fundamental Theorem of Algebra)**.** *A non-constant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.*

From Theorem 2.3 one can deduce that a degree $n \geq 1$ complex polynomial has exactly n roots. Some of these roots though my repeat more than once. For example, $x^2 - 2x + 1$ can be written as $(x-1)^2$ from which one would agree that 1 counts as 2 roots of the polynomials, whatever that means.
Theorem 2.3 wasn't obvious even for the great mathematicians of the time. For example, Bernoulli proposed a counterexample of degree 4. The great Euler proved him wrong in a letter to Goldbach. Euler claimed he has a proof for all degrees $n \leq 6$. A proof for the general case had to wait for Gauss who used trigonometric series in his 1699' proof.
For the Ph.D. students who are reading this, Gauss proved the theorem while being a Ph.D. student. Just saying :) Gauss, being Gauss, subsequently gave 3 other proofs. By now, there are many proofs, non of which is very easy, but you can fit one to a page or two (see Stewart's book). I'm not going to give a proof here. I'm gonna do something even better–I'm gonna show you *why* the theorem is true! The proof sketch is "topological" in nature, i.e., we're going to stretch continuous stuff in a continuous way. Also, I am kind of going to assume that you know about polar presentation.

Proof Sketch. Say * you are looking at a polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$. If $a_0 = 0$ then $x = 0$ is clearly a root of p. So, $a_0 \in \mathbb{C}$ sits somewhere in the complex plane away from the origin. Consider the following thought experiment. Fix a real number $r \geq 0$ and consider the circle of all $x \in \mathbb{C}$ with modulus $|x| = r$. Where does p map this circle to? Well, I don't quite know. But, if r is very large (compared

*When turned into a formal proof, replace with "Let $p(x)$ be..."

to the coefficients and the degree n) then $a_n x^n$ will be the dominating term. If it was the only term then the image of the map of the circle would have been a circle (in fact, n circles on top of each other) with modulus $a_n r^n$. However, there are these other pesky terms which make the actual image look like a wiggly circle. At any rate, if r is large enough the image is almost a circle.
Now comes the punch line. Starting from the huge r you chose, start to decrease it slowly all the way down to 0. If r is chosen large enough, we can make sure that the wiggly shape will contain both a_0 and the origin. However, we know that at the end, when $r = 0$, the wiggly shape will converge to the single point a_0 and, in particular. As everything we do is "continuous" at some point the wiggly shape–the image of p–must pass through the origin. □

2.3 What else is cool about $\mathbb{C}$?

Well, many things. For one, it turns out that $\mathbb{C}$ is very real. I am no expert, but it seems that complex numbers are at the very least most suitable for describing Quantum Mechanics. Mathematicians like complex numbers partially because working with complex functions is much nicer than with real-valued functions. To give some feeling for it, if you're working with a real function and it has an annoying singularity at some point, in $\mathbb{R}$ the function is "broken" into two pieces. Over $\mathbb{C}$ however you can just "go around" the misbehaved point. You can do much more though. For example, you can take a function that is defined somewhere in the complex plane but not in other possibly huge parts of it and, if the function is nice enough, you can extend it to more or less the whole complex plane in a unique way. It is a typical scenario that the new function shed new light on the original, partially defined, function. One fascinating application of such technique is to number theory and in particular to the Riemann Zeta Function. We'll talk a bit about it later in the course.

2.4 What does a Turing Machine think of $\mathbb{C}$?

The Fundamental Theorem of Algebra is extremely useful in theoretical computer science, coding theory, cryptography and what have you. However, computers (or Turing Machines if you must) don't like these infinite precision kind of number systems like $\mathbb{R}$ and $\mathbb{C}$. Even $\mathbb{Q}$ and $\mathbb{Z}$ are not comfortable computing over as when turning to the analysis, one would need to keep track of the size of the computed numbers which, at best, is daunting.
Luckily, there are "finite number systems" which, being finite, avoid these issues. Once can compute over these finite number systems and prove theorems about them.

In particular, The Fundamental Theorem of Algebra more or less holds for these number systems as well–not just over $\mathbb{C}$. The proof, however, as you might expect looks very different as we're working in a very different setting. Soon we will get to these mysterious finite number systems. We will call them *finite fields*.

2.5 Bézout's Theorem

Another very interesting and useful generalization of Theorem 2.3 is obtained by viewing the whole thing geometrically. First, let's work only over $\mathbb{R}$ so it will be easier to draw things in our head. Theorem 2.3 implies that over $\mathbb{R}$, a degree $n \geq 1$ polynomial $p(x)$ has at most n roots. Geometrically, this means that the set of points $C = \{(x, p(x)) \mid x \in \mathbb{R}\}$ that describe the graph of $p(x)$ in the real plane intersects the x-axis $\{(x, 0) \mid x \in \mathbb{R}\}$ in at most n points. You can easily convince yourself that this holds true not only for the x-axis but actually for any line $\{(x, y) \mid ax + by = c\}$ where $a, b, c \in \mathbb{R}$, not all zero, as long as C does not fully contain the line (this reservation with respect to the x-axis is hidden in the hypothesis of Theorem 2.3 that the degree n of p is greater than 1. This takes out the zero polynomial, whose graph is the x-axis, out of the picture.

We call C an *algebraic curve* (or simply a *curve*). Naturally, we say that C has degree n. The curves that correspond to linear equations have degree 1. So, Theorem 2.3 implies that the number of intersection points between a degree n curve and a degree 1 curve in the plane is at most $n \cdot 1$. What about a degree n curve and a degree m curve? You guessed right! The number of intersection points is at most $n \cdot m$. This holds for even more general curves than "just" those of the form $y = p(x)$. You can mix up x, y in anyway you like. For example, $xy - 1 = 0$ is a degree 2 curve.

This remarkable result is called *Bézout's Theorem*. In fact, more is true. If you work over $\mathbb{C}$ and count repeated points of intersections correctly you can almost say that the number of intersection points will be exactly $n \cdot m$. That is not quite true–think of two parallel lines. Turns out, though, that if you are open about changing your geometry from the standard geometry (called affine geometry) to what is called projective geometry, you get precisely $n \cdot m$ points of intersection. The projective plane can be thought of as adding "points at infinity" to the affine plane, in which parallel lines meet. We won't get into this in this course (ad ahead!) but will do so in a followup course on the fascinating subject of Algebraic Geometric codes. One of the goals of this course is to prepare you for the next one.

Let's close with a fun fact. In his original 1770 paper, Bézout didn't correctly account for multiplicities. As the theorem statement was "in the air", one may argue (as some critics have) that the result is neither original nor correct...

References

1. Ian Nicholas Stewart. *Galois theory*. CRC Press, 2015.