Insolvability of the Quintic

An unofficial sequel to An Inquiry-Based Approach to Abstract Algebra by Dana C. Ernst

Joshua Wiscons
California State University, Sacramento

Spring 2019

The most up-to-date version of these notes on can be found on GitHub:

https://github.com/jwiscons/IBL-InsolvabilityOfQuintic

This work is designed to extend An Inquiry-Based Approach to Abstract Algebra by Dana C. Ernst. The presentation of the material is heavily influenced by the book Abstract Algebra: A Concrete Introduction by Robert H. Redfield. Many thanks to both Dana and Bob! I am also indebted to every student in my Modern Algebra 2 class from Fall 2019 at California State University, Sacramento for taking part, willingly or otherwise, in this experiment—thanks to you all!

Contents

Chapter 1

Introduction

This course is a story. A repackaging of a famous story, spanning more-or-less 4000 years, about solving polynomial equations. I hope you like it. I also hope enjoy the beautiful sights along the way, many of which were a long time in the making and most of which are still being heavily researched to this day.

1.1 Prerequisites

A first course in abstract algebra, focusing on the basics of group theory, together with exposure to foundational topics like primes and divisibility, functions and relations, differential calculus, and linear algebra form the core prerequisites. Comfort with basic proof techniques, including induction, is also more-or-less required. The following two free and open-source books developed by Dana C. Ernst will serve you well if you need to review.

- An Inquiry-Based Approach to Abstract Algebra
- An Introduction to Proof via Inquiry-Based Learning

1.2 An Inquiry-Based Approach

This section of the introduction as well as those that follow are (only slightly) modified from the introduction of An Inquiry-Based Approach to Abstract Algebra. The use of "I" below, does indeed refer to me, but mainly just because I also believe the words that Dana originally wrote.

In many courses, math or otherwise, you sit and listen to a lecture. These lectures may be polished and well-delivered. You may have often been lured into believing that the instructor has opened up your head and is pouring knowledge into it. I love lecturing, and I do believe there is value in it, but I also believe that in reality most students do not learn by simply listening. You must be active in the learning process. Likely, each of you have said to yourselves, "Hmmm, I understood this concept when the professor was going over it, but now that I am alone, I am lost." In order to promote a more active

participation in your learning, we will incorporate ideas from an educational philosophy called inquiry-based learning (IBL).

Loosely speaking, IBL is a student-centered method of teaching mathematics that engages students in sense-making activities. Students are given tasks requiring them to solve problems, conjecture, experiment, explore, create, communicate. Rather than showing just facts or a clear, smooth path to a solution, the instructor guides and mentors students via well-crafted problems through an adventure in mathematical discovery. Effective IBL courses encourage deep engagement in rich mathematical activities and provide opportunities to collaborate with peers (either through class presentations or group-oriented work).

Perhaps this is sufficiently vague, but I believe that there are two essential elements to IBL. Students should as much as possible be responsible for:

(1) Guiding the acquisition of knowledge, and

(2) Validating the ideas presented. That is, students should not be looking to the instructor as the sole authority.

For additional information, check out Dana's blog post, What the Heck is IBL?

Much of the course will be devoted to students proving theorems on the board and a significant portion of your grade will be determined by the mathematics you produce. I use the word "produce" because I believe that the best way to learn mathematics is by doing mathematics. Someone cannot master a musical instrument or a martial art by simply watching, and in a similar fashion, you cannot master mathematics by simply watching; you must do mathematics!

Furthermore, it is important to understand that proving theorems is difficult and takes time. You should not expect to complete a single proof in 10 minutes. Sometimes, you might have to stare at the statement for an hour before even understanding how to get started.

In this course, everyone will be required to

- read and interact with course notes on your own;
- write up quality proofs to assigned problems;
- present proofs on the board to the rest of the class;
- participate in discussions centered around a student's presented proof;
- call upon your own prodigious mental faculties to respond in flexible, thoughtful, and creative ways to problems that may seem unfamiliar on first glance.

As the semester progresses, it should become clear to you what the expectations are. This will be new to many of you and there may be some growing pains associated with it.

Lastly, it is highly important to respect learning and to respect other people's ideas. Whether you disagree or agree, please praise and encourage your fellow classmates. Use ideas from others as a starting point rather than something to be judgmental about. Judgement is not the same as being judgmental. Helpfulness, encouragement, and compassion are highly valued.

1.3 Rules of the Game

You should *not* look to resources outside the context of this course for help. That is, you should not be consulting the Internet, other texts, other faculty, or students outside of our course. On the other hand, you may use each other, the course notes, me, and your own intuition.

However, there do exist some hints, collected in Appendix A. Everyone may use the hints, but they were included mostly to support those who need to miss a class and are not able to find a time to meet with their classmates (or me). If you use the hints, please keep in mind that (1) your own learning will significantly benefit from cognitive struggles (independently and with your peers), so don't turn to the hints too early; and (2) the hints are really just some possible ways to get started—they may very well not be the way that makes the most sense to you, so I encourage you to follow you own paths.

In this class, earnest failure outweighs counterfeit success; you need not feel pressure to hunt for solutions outside your own creative and intellectual reserves.

1.4 Structure of the Notes

As you read the notes, you will be required to digest the material in a meaningful way. It is your responsibility to read and understand new definitions and their related concepts. However, you will be supported in this sometimes difficult endeavor. In addition, you will be asked to complete problems aimed at solidifying your understanding of the material. Most importantly, you will be asked to make conjectures, produce counterexamples, and prove theorems.

The items labeled as **Definition** and **Example** (and occasionally **Fact**) are meant to be read and digested. However, the items labeled as **Problem**, **Lemma**, **Theorem**, and **Corollary** require action on your part. Items labeled as **Problem** are sort of a mixed bag. Some Problems are computational in nature and aimed at improving your understanding of a particular concept while others ask you to provide a counterexample for a statement if it is false or to provide a proof if the statement is true. Items with the **Lemma**, **Theorem**, and **Corollary** designation are mathematical facts and the intention is for you to produce a valid proof of the given statement. **Lemma's** are usually stepping stones to the next theorem, though they are often interesting in their own right. **Corollaries** are typically statements that follow quickly from a previous theorem. In general, you should expect corollaries to have very short proofs. However, that doesn't mean that you can't produce a more lengthy yet valid proof of a corollary.

It is important to point out that there are very few examples in the notes. This is intentional. One of the goals of the items labeled as **Problem** is for you to produce the examples.

Lastly, there are many situations where you will want to refer to an earlier definition, problem, lemma, theorem, or corollary. In this case, you should reference the statement by number (or by name if it has one).

Chapter 2

Solving polynomial equations and the main question

The story begins. Can you count all of the times in your career that you've had to find the zeros of a quadratic polynomial? What about a cubic polynomial? A quartic? Likely your answers are decreasing rapidly, and it's also likely that you have only solved cubic and quartic equations in very special situations. Why? Is it just that cubic and quartic equations are difficult to solve or could it be that some are impossible to "solve."

People have been investigating how to solve polynomial equations for about 4000 years. Let's get started.

Problem 2.1. Determine the roots (i.e. zeros) of each of the following. Try to use tools you've accumulated over the years, but you may well need a computer program (e.g. WolframAlpha) for some of them. For each problem, make a note about *how* you found the roots. Try for exact answers, but you can approximate if needed.

(1) $p(x) = x^2 - 5x + 6$

(2) $q(x) = (x-3)^2 - 2$

(3) $r(x) = x^2 + x + 1$

(4) $s(x) = x^3 - 3x - 2$

(5) $f(x) = x^3 - 5$

(6) $g(x) = x^4 - 1$

(7) $a(x) = x^5 - 1$

(8) $b(x) = x^5 + 5x^4 - 5$

Problem 2.2. For each part of Problem 2.1, write down the "smallest" number system needed to express the roots of the given polynomial. Possible answers might be $\mathbb{Z}$ (integers), $\mathbb{Q}$ (rational numbers), $\mathbb{Z}$ together with $\sqrt{3}$, $\mathbb{Q}$ together with $\sqrt{-1}$ and $\sqrt[3]{5}$, etc.

2.1 Solving polynomial equations with formulas

Though you may have solved the first three parts of Problem 2.1 different ways, there was one tool that would have solved them all: the quadratic formula. It will be valuable to (re)discover why it's true.

First, let's slightly simplify things. Notice that α is a root of ax^2+bx+c if and only if α is a root of $x^2+\frac{b}{a}x+\frac{c}{a}$ (assuming $a \neq 0$). This means that an arbitrary quadratic polynomial can always be converted to a quadratic polynomial whose leading coefficient is 1 and in such a way that they have the same roots. Thus, if we have a formula that finds the roots of so-called *monic* quadratic polynomials, we can actually use it to find the roots of *all* quadratic polynomials.

Definition 2.3. A polynomial whose leading coefficient is 1 is called a **monic** polynomial.

Now, let's (re)derive the quadratic formula for monic polynomials. Remember, we are *(re)deriving* it, so *please don't use the quadratic formula in your proof of the next theorem.*

Theorem 2.4. The roots of $p(x) = x^2 + bx + c$ are

$$\frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

Problem 2.5. Describe a number system, as small as possible, that is capable of expressing the roots of *any* quadratic polynomial whose *coefficients are all rational numbers*.

Well, that takes care of quadratic polynomials. What about finding the roots of cubic polynomials? Well, it turns out that there is indeed a cubic formula, though it's decidedly more complicated than the quadratic formula.

A method for deriving a cubic formula, due to Scipione del Ferro and Tartaglia, was published in a book by Cardano in 1545. The starting point is to take a general cubic polynomial and first convert it to a monic polynomial (as we did above) and then convert it to a cubic of the form x^3+px+q, always with the same roots as the original. Then, with work, one arrives at the following formula.

Fact 2.6. The roots of $p(x) = x^3 + px + q$ are

$$\alpha+\beta, \left(-\frac{1}{2}+\frac{\sqrt{-3}}{2}\right)\alpha+\left(-\frac{1}{2}-\frac{\sqrt{-3}}{2}\right)\beta, \left(-\frac{1}{2}-\frac{\sqrt{-3}}{2}\right)\alpha+\left(-\frac{1}{2}+\frac{\sqrt{-3}}{2}\right)\beta$$

where $\alpha = \sqrt[3]{-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}$ and $\beta = \sqrt[3]{-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}$

Problem 2.7. Describe a number system, as small as possible, that is capable of expressing the roots of *any* cubic polynomial of the form $x^3 + px + q$ where $p, q \in \mathbb{Q}$.

Problem 2.8. Use Fact 2.6 to write out the roots of $p(x) = x^3 - 2x - 4$. Are any of the roots integers? Check your answer with WolframAlpha. Does this expose any issues with using the formula?

For more details on solving cubic equations, you can use start with the Wikipedia page about cubic functions. And now... quartic functions? Perhaps you have a guess.

Problem 2.9. Use the internet and/or library to determine if there is a quartic formula, i.e. a formula to find the roots of fourth-degree polynomials. If there is a quartic formula, who are some people that discovered methods to derive it and when did they discover their method?

2.2 The main question(s)

We now turn our attention to finding the roots of polynomials of degree five and higher. Well, what do you think: is there a quintic formula? Actually, there are two questions here: (1) what do we mean by "formula" (e.g. what symbols/functions can we use), and (2) whatever we mean by formula, is there one that works for *all* quintic polynomials?

You investigated two particular quintics in Problem 2.1—what did you find? Did you find exact expressions for the roots? If so, *how* were the roots expressed; that is, what symbols/functions were needed to write out the roots?

Let's first try to tackle the "what do we mean by formula" question. Guided by the quadratic and cubic formulas, let's agree that we are looking for a formula that expresses the roots of an arbitrary quintic polynomial in terms of the coefficients of the polynomial using just the operations of addition, subtraction, multiplication, division, and the extraction of roots (square roots, cube roots,...). This leads to the following intuitive definition, that we will work to sharpen later.

Intuitive Definition 2.10. A polynomial is said to be **solvable by radicals** if every root of the polynomial can be expressed in terms of the coefficients of the polynomial using the operations of addition, subtraction, multiplication, division, and $\sqrt[n]{\ }$ for any positive integer n.

Problem 2.11. Find the roots of $p(x) = x^4 - 2x^2 - 1$. Explain why $p(x)$ is solvable by radicals.

Problem 2.12. Explain why every quadratic polynomial is solvable by radicals.

Returning to quintic polynomials, our main question is as follows.

Main Question. Does there exist a "quintic formula" that expresses the roots of an arbitrary quintic polynomial, in terms of the coefficients of the polynomial, using just rational numbers together with the operations of addition, subtraction, multiplication, division, and the extraction of roots?

Well, if the answer is yes, then it must be that every quintic polynomial is solvable by radicals. *Spoiler Alert!* Our goal (for the rest of the book!) is to prove the following theorem, in a rather elegant way.

Main Theorem. Not every quintic polynomial is solvable by radicals.

Main Corollary. There is *no* "quintic formula" that expresses the roots of an arbitrary quintic polynomial in terms of the coefficients of the polynomial using just the operations of addition, subtraction, multiplication, division, and the extraction of roots.

Boom!

Chapter 3

Fields

In Chapter 2, we explored problems about finding and expressing roots of polynomials, finally arriving at the goal of the course: proving that there are quintic polynomials that are *not* solvable by radicals. This chapter serves two main purposes. First, as we look at roots of polynomials and how they can be expressed, it will be convenient to have a common world (i.e. number system) in which they live. For us, this will be the complex numbers, denoted $\mathbb{C}$, which will be reviewed below. Our work with complex numbers will also supply the necessary language to properly talk about n^{th}-roots. Second, we are still in need of a proper definition of what it means for a polynomial to be "solvable by radicals"; this is where the chapter will finish. But the middle of the chapter is perhaps the most interesting. There, on the way to defining "solvable by radicals", we will be led to abstract the structure of $\mathbb{C}$ (and of $\mathbb{Q}$ and $\mathbb{R}$), arriving at the definition of a *field*.

3.1 Complex Numbers

As mentioned above, we want to work in a world that contains all of the roots of all of the polynomials that we will be studying. Considering the roots of polynomials such as x^2+1, x^2-2, x^2-3, etc., we see that we need to include numbers like $\sqrt{-1}$, $\sqrt{2}$, $\sqrt{3}$, etc., so although there are smaller worlds one could choose, we will opt for the world containing both $\sqrt{-1}$ and $\mathbb{R}$, namely $\mathbb{C}$.

But before we proceed, note that $\sqrt{-1}$ is not really well defined. There are *two* solutions to x^2+1, so when we write $\sqrt{-1}$, we are all agreeing that we mean the same one.

Definition 3.1. Let i (or alternatively $\sqrt{-1}$) denote one particular solution to x^2+1.

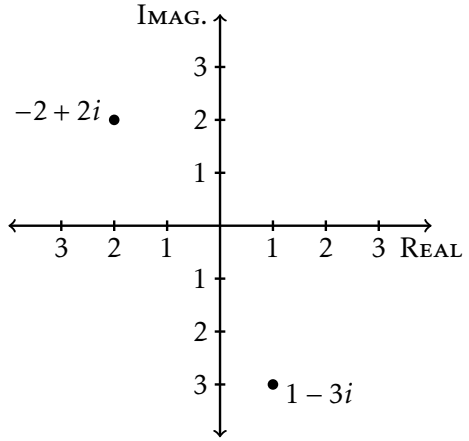
Of course, the previous definition implies that $i^2 = -1$. Using i and $\mathbb{R}$, we now build the complex numbers.

3.1.1 Definition and first principles

Definition 3.2. The **complex numbers** is the set $\mathbb{C} := \{a+bi \mid a,b \in \mathbb{R}\}$. If $z = a+bi$, then a is called the **real part** of z and b is called the **imaginary part** of z.

Note that every complex number $z = a + bi$ is uniquely determined by two numbers: the real and imaginary parts a and b. As such, we often graph complex numbers in the coordinate plane with the x-axis denoting the real part and the y-axis denoting the imaginary part. This will be called the **complex plane**.

Example 3.3. We graph $-2 + 2i$ and $1 - 3i$ below.



We also define some operations on complex numbers.

Definition 3.4. We define the following operations on elements of $\mathbb{C}$.

- **Addition:** $(a + bi) + (c + di) := (a + c) + (b + d)i$
- **Multiplication:** $(a + bi) \cdot (c + di) := (ac - bd) + (ad + bc)i$
- **Complex Conjugation:** $\overline{a + bi} := a - bi$

Notice that in the definition of complex multiplication we are just using the normal distributive law (or FOIL if you like) together with the fact that $i^2 = -1$. Many of the familiar algebraic properties of $\mathbb{R}$ also hold for $\mathbb{C}$, which we will take as a fact.

Fact 3.5. The following are true for $\mathbb{C}$.

- **Addition Laws:** Addition is associative and commutative. There is a unique additive identity, namely $0 = 0 + 0i$, and every number has a unique additive inverse, denoted $-(a + bi)$.
- **Multiplication Laws:** Multiplication is associative and commutative. There is a unique multiplicative identity, namely $1 = 1 + 0i$, and every nonzero number has a unique multiplicative inverse, denote $(a + bi)^{-1}$ or $\frac{1}{a+bi}$.
- **Distributivity Laws:** For all $x, y, z \in \mathbb{C}$, $x(y + z) = xy + xz$ and $(y + z)x = yx + zx$.
- **Conjugation Laws:** For all $x, y \in \mathbb{C}$, $\overline{x + y} = \overline{x} + \overline{y}$ and $\overline{xy} = \overline{xy}$.

Problem 3.6. Thinking of a complex number $z = a+bi$ as a point in the complex plane, describe *geometrically* what happens when $(c + di)$ is added to z. Also, describe *geometrically* how to find $\overline{z}$ from z.

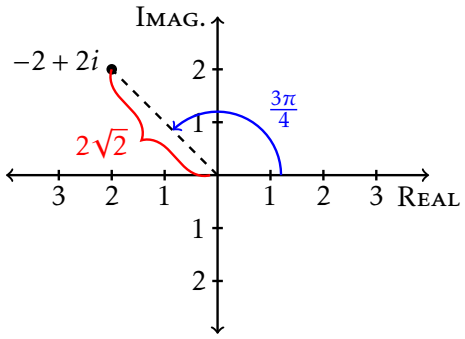
When we plot points, there are different coordinate systems we could use. It turns out that rectangular coordinates are good for adding complex numbers, but polar coordinates are better for multiplication. This lead to the following definition.

Definition 3.7. Let $z = a + bi$.

(1) The **modulus** of z, denoted $|z|$, is the radius of the point (a, b) when written in polar coordinates. Thus, $|z| = \sqrt{a^2 + b^2}$.

(2) The **argument** of z, denoted $\text{Arg}(z)$, is the angle of the point (a, b) when written in polar coordinates. Thus, $\text{Arg}(z)$ is the angle θ *in the appropriate quadrant* such that $0 \leq \theta < 2\pi$ and $\tan\theta = \frac{b}{a}$. The argument of 0 is undefined.

Example 3.8. We have that $|-2 + 2i| = \sqrt{(-2)^2 + 2^2} = 2\sqrt{2}$ and $\text{Arg}(-2 + 2i) = \frac{3\pi}{4}$. (But be careful, $\arctan\left(\frac{2}{-2}\right) = \frac{\pi}{4}$; you must pay attention to which quadrant the number is in.)



Problem 3.9. For each of the following complex numbers,

- write it in the form $a + bi$ (if it is not already),
- plot it in the complex plane,
- find the modulus and argument (if not exact, then a decimal approximation is okay).

(1) $u = -1 - i$

(2) $v = \frac{1}{1+i}$

(3) $w = \frac{(2-i)(1+2i)}{2+3i}$

(4) $z \in \mathbb{C}$ with $|z| = 3$ and $\text{Arg}(z) = \frac{4\pi}{3}$

Theorem 3.10. Let $z \in \mathbb{C}$. If $z \neq 0$, then $z^{-1} = \dfrac{\overline{z}}{|z|^2}$.

The next theorem shows how to find an expression for a complex number given its modulus and argument.

Theorem 3.11. Let $z \in \mathbb{C}$. Then $|z| = r$ and $\text{Arg}(z) = \theta$ if and only if $z = r\cos\theta + ir\sin\theta$ with $0 \leq \theta < 2\pi$.

We now derive some properties of multiplication. The first is quite useful and illustrates how multiplication is rather easy to deal with when numbers are in "polar form".

Theorem 3.12. If $z_1 = r_1\cos\theta_1 + ir_1\sin\theta_1$ and $z_2 = r_2\cos\theta_2 + ir_2\sin\theta_2$, then

$$z_1z_2 = r_1r_2\cos(\theta_1+\theta_2) + ir_1r_2\sin(\theta_1+\theta_2).$$

Corollary 3.13. If $z_1, z_2 \in \mathbb{C}$, then $|z_1z_2|$ is equal to $|z_1||z_2|$ and $\mathrm{Arg}(z_1z_2)$ is equivalent to $\mathrm{Arg}(z_1) + \mathrm{Arg}(z_2)$ modulo 2π.

Corollary 3.14 (De Moivre's formula)**.** For each positive $n \in \mathbb{Z}$,

$$(r\cos(\theta) + ir\sin(\theta))^n = r^n\cos(n\theta) + ir^n\sin(n\theta).$$

3.1.2 Roots of unity

We now arrive at an *extremely important* definition.

Definition 3.15. For each positive $n \in \mathbb{Z}$, define

$$\zeta_n := \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right).$$

Thus, ζ_n (read as "zeta n") is the unique number with magnitude 1 and argument $\frac{2\pi}{n}$.

Problem 3.16. Plot each of the following in the same complex plane: ζ_2, ζ_3, ζ_4, ζ_5.

Problem 3.17. Plot each of the following in the same complex plane: ζ_6, $(\zeta_6)^2$, $(\zeta_6)^3$, $(\zeta_6)^4$, $(\zeta_6)^5$, $(\zeta_6)^6$.

Problem 3.18. Write $\overline{\zeta_8}$ as a power of ζ_8. Conjecture and prove a formula that expresses $\overline{(\zeta_n)^k}$ as a power of ζ_n, but with no bar on top.

We now turn our attention back to solving polynomial equations, focusing on those of the form $x^n - a$.

Definition 3.19. Let $a \in \mathbb{C}$. A number $z \in \mathbb{C}$ is called an n^{th} **root of** a if $z^n = a$. In other words, the n^{th} roots of a are the roots of the polynomial $x^n - a$. The n^{th} roots of 1 are also called n^{th} **roots of unity**.

Problem 3.20. Find a 4^{th} root of each of the following: ζ_3 and $-1 + i\sqrt{3}$.

Theorem 3.21. For each non-negative $k \in \mathbb{Z}$, $(\zeta_n)^k$ is an n^{th} root of 1.

Lemma 3.22. If z is an n^{th} root of 1, then $z = (\zeta_n)^k$ for some non-negative $k \in \mathbb{Z}$.

Lemma 3.23. For each non-negative $k \in \mathbb{Z}$, $(\zeta_n)^k = (\zeta_n)^m$ for some $0 \le m \le n-1$.

Theorem 3.24. The set

$$\{1, \zeta_n, (\zeta_n)^2, \ldots, (\zeta_n)^{n-1}\}$$

is the set of *all* n^{th} roots of unity. Thus, there are n distinct n^{th} roots of unity.

Lemma 3.25. Let $a \in \mathbb{C}$ be nonzero, and let b be any one particular n^{th} root of a. Then z is an n^{th} root of a if and only if $\frac{z}{b}$ is an n^{th} root of 1.

Theorem 3.26. Let $a \in \mathbb{C}$ be nonzero, and let b be any one particular n^{th} root of a. The set

$$\{b, b\zeta_n, b(\zeta_n)^2, \ldots, b(\zeta_n)^{n-1}\}$$

is the set of *all* n^{th} roots of a. Thus, there are n distinct n^{th} roots of a.

Problem 3.27. Find *all* 4^{th} roots of each of the following: ζ_3 and $-1 + i\sqrt{3}$.

3.1.3 Roots of polynomials over $\mathbb{R}$ and $\mathbb{C}$

We conclude this section with a couple of general results about roots of polynomials.

Theorem 3.28. Suppose that $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ with all $a_i \in \mathbb{R}$. If z is a root of $p(x)$, then $\overline{z}$ is also a root of $p(x)$.

In words, the previous theorem says that if a polynomial has coefficients in $\mathbb{R}$, then the set of roots is "closed under complex conjugation." We end with an extremely important theorem, which will be quite useful for us. However, since its proof is not our main goal (and since it requires sophisticated techniques), we will take it as fact.

Fact 3.29 (Fundamental Theorem of Algebra)**.** If $p(x)$ is a non-constant polynomial with all coefficients in $\mathbb{C}$, then $p(x)$ has a root in $\mathbb{C}$.

In fact, we will see that this implies that *all* roots of such a $p(x)$ lie in $\mathbb{C}$, so in our of study polynomials (often with all coefficients even in $\mathbb{Q}$), $\mathbb{C}$ serves as a uniform world in which we can study the roots.

3.2 An aside: the quaternions

Our construction of the complex numbers creates a structure that contains the real numbers and possesses some nice properties not enjoyed by the real numbers, e.g. every non-constant polynomial with complex coefficients has a complex root. This raises the question: could we further extend the complex numbers to an even larger structure?

Concisely, we built the complex numbers as the set $\mathbb{C} = \mathbb{R} + \mathbb{R}i$ together with the operations of addition and multiplication, which were defined in a natural way from the key identity that $i^2 = -1$. Here, we briefly explore what happens if we build a larger structure in a similar way: $\mathbb{H} = \mathbb{C} + \mathbb{C}j$ where, again, $j^2 = -1$.

Following this path, we formally arrive at $\mathbb{H} = \mathbb{C} + \mathbb{C}j = (\mathbb{R} + \mathbb{R}i) + (\mathbb{R} + \mathbb{R}i)j$, and any definition we give for multiplication of two elements of $\mathbb{H}$ must first define how to multiply i and j (or rather, what properties ij should have). If we set $k = ij$, it turns out that a good route to follow is to decide that k also has the property that it squares to 1, i.e. $k^2 = -1$. There is another important choice one is "forced" to make, namely that $ji = -k$.

Definition 3.30. The **quaternions** are the elements of $\mathbb{H} := \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}$, where $i^2 = j^2 = k^2 = -1$. We also define the following operations on elements of $\mathbb{H}$.

- **Addition:** $(a_1 + b_1 i + c_1 j + d_1 k) + (a_2 + b_2 i + c_2 j + d_2 k) := (a_1 + a_2) + (b_1 + b_2)i + (c_1 + c_2)j + (d_1 + d_2)k$
- **Multiplication:** use the usual distributive laws together with the identities:

$$\begin{array}{lll} ij = k, & jk = i, & ki = j, \\ ji = -k, & kj = -i, & ik = -j. \end{array}$$

- **Conjugation:** $\overline{a + bi + cj + dk} := a - bi - cj - dk$

Indeed, $\mathbb{H}$ extends the complex numbers, and we have the following containments: $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \subset \mathbb{H}$. It turns out that $\mathbb{H}$ satisfies nearly all of the common algebraic properties of $\mathbb{R}$ and $\mathbb{C}$, with one notable exception, which is highlighted in bold below.

Fact 3.31. The following are true for $\mathbb{H}$.

- **Addition Laws:** Addition is associative and commutative. There is a unique additive identity, namely $0 = 0 + 0i + 0j + 0k$, and every number has a unique additive inverse.
- **Multiplication Laws:** Multiplication is associative but **noncommutative**. There is a unique multiplicative identity, namely $1 = 1+0i+0j+0k$, and every nonzero number has a unique multiplicative inverse.
- **Distributivity Laws:** For all $x, y, z \in \mathbb{H}$, $x(y+z) = xy + xz$ and $(y+z)x = yx + zx$.
- **Conjugation Laws:** For all $x, y \in \mathbb{H}$, $\overline{x+y} = \overline{x} + \overline{y}$ and $\overline{xy} = \overline{xy}$.

We can also define the modulus of a quaternion, analogous to how we defined it for a complex number.

Definition 3.32. The **modulus** of $h = a+bi+cj+dk$, denoted $|h|$, is $|h| = \sqrt{a^2+b^2+c^2+d^2}$.

Problem 3.33. Write each of the following quaternions in the form $a + bi + cj + dk$, and find its modulus.

(1) $h = (2i + 4k)(7 - 3j + k)$

(2) k^{-1}

(3) $w = (i + j)^{-1}$

Theorem 3.34. Let $h \in \mathbb{H}$. If $h \neq 0$, then $h^{-1} = \dfrac{\overline{h}}{|h|^2}$.

And as in the complex numbers, the modulus function is multiplicative—we will take this as fact.

Fact 3.35. If $h_1, h_2 \in \mathbb{H}$, then $|h_1 h_2| = |h_1||h_2|$.

We conclude this section by looking at multiplication of quaternions a little closer. As we do, we return to the mathematical notion of a *group*. Please feel free to look over old notes or other books to review the basics. As mentioned in the introduction, our main reference for groups will be An Inquiry-Based Approach to Abstract Algebra.

Problem 3.36. Let G be the subset of $\mathbb{H}$ defined as $G := \{\pm 1, \pm i, \pm j, \pm k\}$. Show that G, together with the operation of quaternion multiplication, is a nonabelian group. If you have encountered this group before, what name (or symbol) did you know it by?

Problem 3.37. Let S be the subset of $\mathbb{H}$ consisting of all quaternions with modulus equal to 1, i.e. $S := \{h \in \mathbb{H} \mid |h| = 1\}$. Show that S, together with the operation of quaternion multiplication, is an infinite, nonabelian group.

It turns out that the group S from the previous problem is isomorphic to the group SU(2) (one the the so-called special unitary groups), which is quite important in theoretical physics. If you want to learn more, you can start on Wikipedia.

3.3 Abstract fields

Notice that $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ satisfy many common algebraic properties with respect to addition and multiplication. Of course, $\mathbb{H}$ does too, though it lacks commutativity of multiplication. When objects have common properties, it can be extremely valuable to abstract those properties and study them once and for all (as opposed to trying to prove things about each individual structure). This is where we are headed, but first we highlight some related structures (again with algebraic properties similar to $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$) that help to connect this work to our main goal of expressing roots of polynomials.

Problem 3.38. Let $p(x) = x^2 + 3x + 1$. Find the roots of $p(x)$, and show that each root can be written in the form $a + b\sqrt{5}$ with $a, b \in \mathbb{Q}$.

Problem 3.39. Let $S = \{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\}$.

(1) Show that S is closed under addition; that is, show that for all $x, y \in S$, $x + y \in S$.

(2) Show that S is closed under multiplication; that is, show that for all $x, y \in S$, $xy \in S$.

(3) Use that $S \subset \mathbb{R}$ to explain why both addition and multiplication of elements of S are associative and commutative and why multiplication distributes over addition.

Problem 3.40. Let $S = \{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\}$. Prove or disprove: if $x \in S$ and $x \neq 0$, then x has a multiplicative inverse *in* S (i.e. there is a $y \in S$ such that $xy = 1$).

3.3.1 Definition

We now abstract the common properties of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ (and also S from Problem 3.39), arriving at the definition of a field.

Definition 3.41. A **field** is a structure $(F, +, \cdot)$ consisting of a set F, containing at least two elements, together with two binary operations $+$ and $\cdot$ (which we call *addition* and *multiplication*) such that for some elements $0, 1 \in F$ the following axioms hold.

- **Addition Axioms:** Addition is associative and commutative; the element 0 is an additive identity; every $x \in F$ has an additive inverse with respect to 0, denoted $-x$.
- **Multiplication Axioms:** Multiplication is associative and commutative; the element 1 is a multiplicative identity; every $x \in F \setminus \{0\}$ has a multiplicative inverse with respect to 1, denoted x^{-1}.
- **Distributivity Axioms:** For all $x, y, z \in F$, $x(y + z) = xy + xz$ and $(y + z)x = yx + zx$.

Recall that "0 is an additive identity" means that "for all $x \in F$, $0 + x = x + 0 = x$," and "$x \in F$ has an additive inverse with respect to 0" means that "there exists some $y \in F$ such that $x + y = y + x = 0$." The meanings of multiplicative identities and inverses are similar to those for addition. Also, recall that $F \setminus \{0\}$ denotes the set obtained by removing the element 0 from F. We introduce some notation for this.

Definition 3.42. If F is a field, then $F \setminus \{0\}$ is denoted by F^*, i.e. F^* is the set of nonzero elements of F.

Using the language of groups, fields can be concisely defined as structures of the form $(F, +, \cdot)$ such that $(F, +)$ is an abelian group with identity 0, $(F^*, \cdot)$ is an abelian group with identity 1, and multiplication distributes over addition.

Now, as with any new definition, we look for examples and basic properties.

3.3.2 Examples and non-examples

It is not hard to verify that $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and S from Problem 3.39 are all fields (with their usual definitions of addition and multiplication). Let's search for more examples and non-examples.

Problem 3.43. Explain why $\mathbb{Z}$ is not a field.

Problem 3.44. Determine if each of the following is a field. If it is a field, identify an additive and multiplicative identity; if it is not a field, explain why not.

(1) $(F, +, \cdot)$ where $F = \{a, b, c\}$ and $+$ and $\cdot$ are defined as follows:

$+$	a	b	c
a	b	c	a
b	c	a	b
c	a	b	c

$\cdot$	a	b	c
a	a	b	c
b	b	a	c
c	c	c	c

(2) $(F, +, \cdot)$ where $F = \{0, 1, 2, 3\}$ and $+$ and $\cdot$ are defined as follows:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

(3) $(F, +, \cdot)$ where $F = \{0, 1, 2, 3\}$ and $+$ and $\cdot$ are defined as follows:

$+$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

Problem 3.45. Look back at Problem 3.44. For those that are fields, determine which familiar group each of $(F, +)$ and $(F^*, \cdot)$ is isomorphic to.

To find more examples of fields, Problem 3.44 hints at the fact we may want to look back to modular arithmetic. Following An Inquiry-Based Approach to Abstract Algebra, we define the structures $(\mathbb{Z}_n, +_n, \cdot_n)$ as follows.

Definition 3.46. Let n be a positive integer. The structure $(\mathbb{Z}_n, +_n, \cdot_n)$ consists of the set $\mathbb{Z}_n := \{0, 1, 2, \ldots, n-1\}$ together with the operations $+_n$ and $\cdot_n$ defined as follows.

- **Addition:** $x +_n y$ is the least non-negative number congruent to $x + y$ modulo n.
- **Multiplication:** $x \cdot_n y$ is the least non-negative number congruent to $x \cdot y$ modulo n.

We often refer to the entire structure $(\mathbb{Z}_n, +_n, \cdot_n)$ as simply $\mathbb{Z}_n$. Also, when the context is clear, we may write $+$ and $\cdot$ in place of $+_n$ and $\cdot_n$.

So, in $\mathbb{Z}_5$, we write equations like $3+6=4$, since $3+6=9$ and 9 is congruent to 4 when working modulo 5. If needed, we can highlight that we are working modulo 5 by writing $3+6 \equiv_5 4$. And with respect to multiplication in $\mathbb{Z}_5$, we have equations like $2 \cdot 3 = 1$, which implies that 3 is a multiplicative inverse of 2 (and vice versa) in $\mathbb{Z}_5$.

Problem 3.47. Show that $\mathbb{Z}_5$ is a field but $\mathbb{Z}_6$ is not.

Problem 3.48. Make a conjecture as to when $\mathbb{Z}_n$ is a field and when it is not. That is, try to fill in the blank: "$\mathbb{Z}_n$ is a field provided ____(something about n)____." What evidence do you have to support this?

3.3.3 Basic properties

Let's now explore some basic properties of fields that follow from the definition. We list some of these as facts since they follow directly from basic group theory, remembering that, as observed above, $(F, +)$ and $(F^*, \cdot)$ are both groups.

From now on, when we write "let F be a field," we tacitly mean "let $(F, +, \cdot)$ be a field."

Fact 3.49. Let F be a field.

(1) The additive identity and the multiplicative identity are both unique.

(2) Additive inverses and multiplicative inverses are unique.

Theorem 3.50. Let F be a field.

(1) For all $x \in F$, $x \cdot 0 = 0$.

(2) For all $x, y \in F$, $(-x)y = -(xy)$ and $x(-y) = -(xy)$.

(3) For all $x \in F^*$, $-x \in F^*$ and $(-x)^{-1} = -(x^{-1})$.

(4) For all $x, y \in F$, if $xy = 0$, then $x = 0$ or $y = 0$.

(5) The additive and multiplicative identities are different, i.e. $0 \neq 1$.

3.3.4 Another example

We now return to the conjecture you made in Problem 3.48. Combining the next theorem with Theorem 3.50, we see that $\mathbb{Z}_n$ has no hope to be a field unless n is prime.

Theorem 3.51. Let n be a positive integer. If n is not prime, then there exist $a, b \in (\mathbb{Z}_n)^*$ such that $ab = 0$ in $\mathbb{Z}_n$.

And now we completely answer the question. As you explore the next theorem, you can use properties of modular arithmetic that you know from before. For example, you can take for granted that addition and multiplication are both associative and commutative. The crux is in showing that every nonzero element has a multiplicative inverse when n is prime. There are many ways to approach this; one way uses Bézout's lemma from basic number theory. Even if you don't use it now, it's a useful fact to remember.

Fact 3.52 (Bézout's lemma)**.** If $a, b \in \mathbb{Z}$, then there exist $k, l \in \mathbb{Z}$ such that $ka + lb = \gcd(a, b)$.

Theorem 3.53. Let n be a positive integer. Then $\mathbb{Z}_n$ is a field if and only if n is prime.

3.4 Subfields and extension fields

Just as with groups and subgroups, the notion of a subfield is extremely important. Analyzing the subfields of a field F can often yield a better understanding of the whole field F, and vice versa. Also, this will allow us to generate more examples of fields.

Definition 3.54. Let $(E, +, \cdot)$ be a field, and let F be a subset of E. Then F is a **subfield** of E if F is a field in its own right with respect to operations $+$ and $\cdot$ *inherited from* E. When F is a subfield of E, we call E an **extension field** of F.

When checking if a subset of a field is a subfield, it turns out that the subset will automatically satisfy many of the field axioms, leaving only a handful of things to verify.

Theorem 3.55. Let E be a field, and let $F \subseteq E$. Then F is a subfield of E if and only if

(1) F contains at least 2 elements;

(2) for all $x, y \in F$, $x + y \in F$ and $xy \in F$;

(3) for all $x \in F$, $-x \in F$; and

(4) for all $x \in F^*$, $x^{-1} \in F$.

The second item in the above theorem is stating that F is closed under the addition and multiplication inherited from E. The last two items could be read as F being closed under additive and multiplicative inverses.

Theorem 3.56. If F is a subfield of E, then F contains the additive and multiplicative identities of E (namely 0 and 1).

It is not difficult to check that $\mathbb{Q}$ and $\mathbb{R}$ are both subfields of $\mathbb{C}$; S from Problem 3.39 is also a subfield of $\mathbb{C}$ (and of $\mathbb{R}$). Let's look for more that are similar to S.

Problem 3.57. Determine which of the following are subfields of $\mathbb{C}$.

(1) $T_1 = \{a + bi \mid a, b \in \mathbb{Q}\}$

(2) $T_2 = \{a + bi \mid a, b \in \mathbb{Z}\}$

(3) $T_3 = \{a + b\alpha \mid a, b \in \mathbb{Q}\}$ where $\alpha = \sqrt{2} + i$

3.4.1 Generating fields

Paralleling the theory of groups, we now investigate how to generate subfields from subsets of elements. We first need a *definition* of "the subfield generated by a set of elements"; it is essentially the same as for all algebraic structures: take the intersection of all subfields containing the subset.

Theorem 3.58. Suppose S is a subset of a field E. Let K be the intersection of all subfields of E that contain S; that is

$$K := \bigcap \{F \mid F \text{ is a subfield of } E \text{ and } S \subseteq F\}.$$

Then

(1) K is a subfield of E that contains S, and

(2) if F is any subfield of E that contains S, then F also contains K.

In words, the previous theorem says that K is the "smallest" subfield of E containing S, so K is the correct candidate for the subfield generated by S.

Definition 3.59. Suppose S is a subset of a field E. The **subfield of E generated by S**, denoted $\langle S \rangle_{\text{FIELD}}$, is defined to be the intersection of all subfields of E that contain S.

In symbols, $S \subseteq \langle S \rangle_{\text{FIELD}} \subseteq E$, and if F is any subfield of E, then $S \subseteq F \implies \langle S \rangle_{\text{FIELD}} \subseteq F$.

Example 3.60. Let's explore $\langle 1 \rangle_{\text{FIELD}}$ in the field $\mathbb{C}$. By definition, $\langle 1 \rangle_{\text{FIELD}}$ is the intersection of all subfields of $\mathbb{C}$ that contain 1.

Let F be an arbitrary subfield of $\mathbb{C}$ containing 1. By Theorem 3.56, every subfield of $\mathbb{C}$ contains 0 and 1, so F must contain 0 (in addition to 1). Further, F must contain $1 + 1$, $1 + 1 + 1$, etc., because F is closed under addition. So, by induction, F contains the positive integers and 0. Then, since F is closed under additive inverses, F also contains the additive inverse of each positive integer, so in total, we now see that F contains $\mathbb{Z}$. Continuing on, F is closed under multiplicative inverses, so F also contains the multiplicative inverse of every nonzero integer. Thus, $\mathbb{Q} \subseteq F$.

Since F was an *arbitrary* subfield of $\mathbb{C}$ containing 1, everything we said above is true for *every* subfield of $\mathbb{C}$ containing 1; thus it is also true for the intersection of them. Hence $\mathbb{Q} \subseteq \langle 1 \rangle_{\text{FIELD}}$. Now we have $\{1\} \subset \mathbb{Q} \subseteq \langle 1 \rangle_{\text{FIELD}}$, so as $\mathbb{Q}$ is a subfield and $\langle 1 \rangle_{\text{FIELD}}$ is the *smallest* subfield containing 1, it must be that $\mathbb{Q} = \langle 1 \rangle_{\text{FIELD}}$.

Theorem 3.61. If $S \subseteq \mathbb{C}$, then $\mathbb{Q} \subseteq \langle S \rangle_{\text{FIELD}}$.

Problem 3.62. The field defined in Problem 3.44(3) is sometimes denoted $\mathbb{F}_4$. Determine $\langle 1 \rangle_{\text{FIELD}}$ in the field $\mathbb{F}_4$.

Most of the time, we will want to generate fields by adding some elements to an existing field, and we have special notation for this.

Notation 3.63. Let F be a subfield of E, and let $r_1, r_2, \ldots, r_n \in E$. The subfield of E generated by $F \cup \{r_1, r_2, \ldots, r_n\}$ is denoted $F(r_1, r_2, \ldots, r_n)$. In other words,

$$F(r_1, r_2, \ldots, r_n) := \langle F \cup \{r_1, r_2, \ldots, r_n\} \rangle_{\text{FIELD}}.$$

We read $F(r_1, r_2, \ldots, r_n)$ as "F adjoin $r_1, r_2, \ldots, r_n$"; it is the smallest field extension of F that contains $r_1, r_2, \ldots, r_n$.

In the following problems, we are working with subfields of $\mathbb{C}$, even if we don't say it explicitly. Thus, by Theorem 3.61, we are working with field extensions of $\mathbb{Q}$.

Problem 3.64. In Problem 3.57(1), we saw that $\{a + bi \mid a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$. Show that $\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$.

Problem 3.65. Show that $\mathbb{R}(i) = \mathbb{C}$.

Problem 3.66. We saw previously that $\{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$. Find some $z \in \mathbb{C}$, such that $\mathbb{Q}(z) = \{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\}$, and prove that your choice for z works. Do you think there is only one choice for z or might others work?

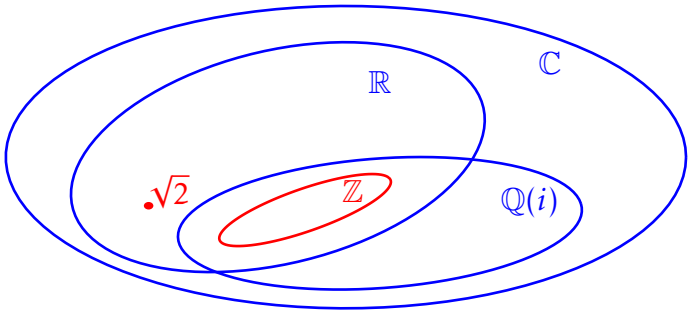
Problem 3.67. Let $\alpha = \sqrt{2} + i$. Show $\{a + b\alpha \mid a, b \in \mathbb{Q}\} \subset \mathbb{Q}(\alpha)$, but $\{a + b\alpha \mid a, b \in \mathbb{Q}\} \neq \mathbb{Q}(\alpha)$.

Theorem 3.68. Let F, L be subfields of E, and let $r_1, r_2, \ldots, r_n \in E$. Then $F(r_1, r_2, \ldots, r_n) \subseteq L$ if and only if $F \subseteq L$ and $r_1, r_2, \ldots, r_n \in L$.

Problem 3.69. Show that $\mathbb{Q}\left(3 - \sqrt{2}, 5 + i\right) = \mathbb{Q}\left(\sqrt{2}, i\right)$.

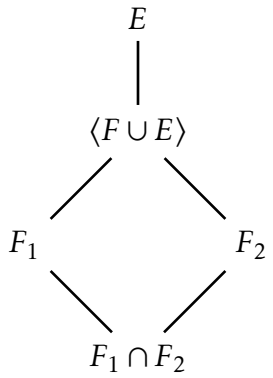
Problem 3.70. Complete the diagram below to illustrate how each of the following sets intersect and where each element is located. Each set that is a field should be drawn in blue; each set that is not a field should be drawn in red. Elements should be illustrated by a dot and then labeled by the name of the element. Some have already been done.

$$\mathbb{C}, \mathbb{R}, \mathbb{Q}, \mathbb{Z}, 0, 1, \sqrt{2}, i, i\sqrt{2}, \sqrt{2} + i, \mathbb{Q}\left(\sqrt{2}\right), \mathbb{Q}(i), \mathbb{Q}\left(i\sqrt{2}\right) \mathbb{Q}\left(\sqrt{2}, i\right), \{a + bi \mid a, b \in \mathbb{Z}\}$$

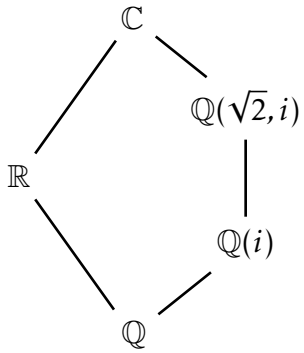


Problem 3.71. Conjecture where $\mathbb{Q}\left(\sqrt{2}+i\right)$ would be in the previous diagram.

Suppose that F_1 and F_2 are subfields of E. Theorem 3.58 tells us that $F_1 \cap F_2$ is again a subfield, and it is the largest subfield contained in both F_1 and F_2. The same theorem, together with Definition 3.59, also tells us that $\langle F_1 \cup F_2 \rangle_{\text{FIELD}}$ is a subfield, and it is the smallest subfield containing both $F_1$and F_2. This implies that the set of all subfields of E forms a lattice. Lattices will not be defined here, but feel free to look them up on your own. We will, however, be interested in illustrating these relationships with a diagram. The situation for F_1 and F_2 described above would be drawn as follows.



For a concrete example, let's draw the portion of the subfield lattice of $\mathbb{C}$ containing $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Q}(i)$, and $\mathbb{Q}\left(\sqrt{2}, i\right)$; this uses some of what you discovered in Problem 3.70.



Problem 3.72. Draw the portion of the subfield lattice of $\mathbb{C}$ that contains the following fields: $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}\left(\sqrt{2}\right)$, $\mathbb{Q}(i)$, $\mathbb{Q}\left(i\sqrt{2}\right)$, and $\mathbb{Q}\left(\sqrt{2}, i\right)$.

Problem 3.73. Draw the portion of the subfield lattice of $\mathbb{C}$ that contains the following fields: $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Q}(\zeta_4)$, $\mathbb{Q}(\zeta_8)$, and $\mathbb{Q}(\zeta_{16})$.

Chapter 4

Solvability by Radicals

Our overarching goal, as laid out in Chapter 2, is to find a polynomial whose roots can **not** be expressed in terms of the coefficients of the polynomial using just the operations of addition, subtraction, multiplication, division, and the extraction of roots. Or, in other words, we are searching for a polynomial that is **not** solvable by radicals, a term that we have only defined informally so far. Laying out a formal definition of solvability by radicals (and trying to wrap our head around it) is the main goal of this chapter.

4.1 Radical extensions

The notion of "solvable by radicals" is about how we may express the roots of a polynomial. We start by formalizing the notion that "a number can be expressed in terms of other numbers using just the operations of addition, subtraction, multiplication, division, and the extraction of roots." In the next section, we apply this to roots of polynomials.

Now, when we define what it means for a number to be built using the various operations listed above, we need to capture the possibility that we may need "iterated roots" to express a number. For example, consider

$$\alpha = \sqrt{2} + \sqrt[3]{-1 + \sqrt{2}}.$$

To see that α can be expressed using addition, subtraction, multiplication, division, and the extraction of roots, we first note that the number $\beta = -1 + \sqrt{2}$ can be built using addition and a square root; we then arrive at α by taking a cube root of β and adding $\sqrt{2}$.

Let's begin to formalize this by introducing fields. Our observations above imply that α can be built using field operations from $\sqrt[3]{\beta}$ and $\sqrt{2}$, and β in turn can be built using field operations from $\sqrt{2}$. Thus, $\alpha \in \mathbb{Q}\left(\sqrt{2}, \sqrt[3]{\beta}\right)$ and $\beta \in \mathbb{Q}\left(\sqrt{2}\right)$. The lattice looks like this.

$$\begin{array}{rl} \alpha \in & \mathbb{Q}\left(\sqrt{2}, \sqrt[3]{\beta}\right) \\ & \Big| \\ \beta \in & \mathbb{Q}\left(\sqrt{2}\right) \\ & \Big| \\ & \mathbb{Q} \end{array}$$

Now, when we talk about extracting roots, we must be careful to avoid ambiguous (not well-defined) notation. For this reason, we usually adopt the point of view of Definition 3.19 where z being an n^{th} root of a means that $z^n = a$ (as opposed to $z = \sqrt[n]{a}$). That said, we do still occasionally use the root symbol when there is no ambiguity. For example, $\sqrt[3]{5}$ and $\sqrt{-1}$ are well-defined: the first is the one and only *real* solution to $x^3 = 5$, and the second is i (which we made a choice about long ago). However, $\sqrt[4]{-1 + i\sqrt{3}}$ is *not* well-defined, as there are 4 equally good choices.

Definition 4.1. We say K is a **radical extension** of a field F if there exist nonzero elements $r_1, r_2, \ldots, r_m \in K$ and positive integers $n_1, n_2 \ldots, n_m$ such that $K = F(r_1, r_2, \ldots, r_m)$, and

$$\begin{aligned} r_1^{n_1} &\in F, \\ r_2^{n_2} &\in F(r_1), \\ r_3^{n_3} &\in F(r_1, r_2), \\ &\vdots \\ r_k^{n_k} &\in F(r_1, \ldots, r_{k-1}). \end{aligned}$$

The definition expresses that each r_i is an n_i^{th}-root of some element in $F(r_1, \ldots, r_{i-1})$, so K may be thought of as being built by iteratively adding in n^{th}-roots of elements. The picture is something like this:

$$\begin{array}{rl} & K = F(r_1, r_2, \ldots, r_m) \\ & \Big| \\ r_m^{n_m} \in & F(r_1, r_2, \ldots, r_{m-1}) \\ & \Big| \\ & \vdots \\ & \Big| \\ r_3^{n_3} \in & F(r_1, r_2) \\ & \Big| \\ r_2^{n_2} \in & F(r_1) \\ & \Big| \\ r_1^{n_1} \in & F \end{array}$$

Example 4.2. Let's show that $K = \mathbb{Q}\left(\sqrt{2}, \sqrt[3]{-1+\sqrt{2}}\right)$ is a radical extension of $\mathbb{Q}$.

If we let $r_1 = \sqrt{2}$ and $r_2 = \sqrt[3]{-1+\sqrt{2}}$, then we see that

- $K = \mathbb{Q}(r_1, r_2)$;
- $r_1^2 = 2 \in \mathbb{Q}$;
- $r_2^3 = -1 + \sqrt{2} \in \mathbb{Q}\left(\sqrt{2}\right) = \mathbb{Q}(r_1)$.

This shows that K is a radical extension of $\mathbb{Q}$ using $r_1 = \sqrt{2}$, $n_1 = 2$, $r_2 = \sqrt[3]{-1+\sqrt{2}}$, and $n_2 = 3$ in the definition of a radical extension. The picture is like this:

$$\begin{array}{ccc} K = \mathbb{Q}\left(\sqrt{2}, \sqrt[3]{-1+\sqrt{2}}\right) & & \\ & | & \\ \left(\sqrt[3]{-1+\sqrt{2}}\right)^3 \in & \mathbb{Q}\left(\sqrt{2}\right) & \\ & | & \\ \left(\sqrt{2}\right)^2 \in & \mathbb{Q} & \end{array}$$

Problem 4.3. Show that $\mathbb{Q}\left(\sqrt{3}, \zeta_5, \sqrt[4]{1-\sqrt{3}}\right)$ is a radical extension of $\mathbb{Q}$. What are you using for r_1, n_1, r_2, n_2, and r_3, n_3 when applying the definition?

Problem 4.4. Show that $\mathbb{C}$ is a radical extension of $\mathbb{R}$.

4.2 Solvability by radicals: the definition

Making precise what it means to express a number using addition, subtraction, multiplication, division, and the extraction of roots was the crux of defining solvability by radicals. However, we are aiming to define what it means to express the roots of a polynomial *in terms of the coefficients* using these operations. Let's establish some notation that allows us to highlight where the coefficients of a polynomial live.

Definition 4.5. Let F be a field. Then $F[x]$ is the set of all polynomials whose coefficients lie in F. This is read "F adjoin x."

For example, consider $a(x) = x^3 - ix^2 - 0.5$. Then $a(x) \notin \mathbb{Q}[x]$, because $i \notin \mathbb{Q}$; however, $a(x) \in \mathbb{C}[x]$ (and, in fact, $a(x) \in \mathbb{Q}(i)[x]$).

Problem 4.6. Give examples of polynomials $a(x)$, $b(x)$, and $c(x)$ such that

(1) $a(x) \in \mathbb{Q}[x]$,

(2) $b(x) \in \mathbb{R}[x]$ but $b(x) \notin \mathbb{Q}[x]$, and

(3) $c(x) \in \mathbb{C}[x]$ but $c(x) \notin \mathbb{R}[x]$.

We now, finally, write down one of our main definitions.

Definition 4.7. Let F be a field, and let $p(x) \in F[x]$. We say that $p(x)$ is **solvable by radicals** over F if all of the roots of $p(x)$ are contained in some radical extension of F.

Problem 4.8. Let $p(x) = x^2 + 3x + 1$. Show that all roots of $p(x)$ lie in $\mathbb{Q}\left(\sqrt{5}\right)$. Use this to explain why $p(x)$ is solvable by radicals over $\mathbb{Q}$.

Problem 4.9. Let $p(x) = x^4 + 2x^2 + 5$. Show that all four roots of $p(x)$ lie in $\mathbb{Q}(i, r, s)$ for some r and s such that $r^2 = -1 - 2i$ and $s^2 = -1 + 2i$. Use this to explain why $p(x)$ is solvable by radicals over $\mathbb{Q}$.

Problem 4.10. Let $p(x) = x^3 - 2$. Use Theorem 3.26 to write out all complex roots of $p(x)$, and then show that $p(x)$ is solvable by radicals over $\mathbb{Q}$.

Theorem 4.11. For each positive $n \in \mathbb{Z}$, $x^n - 1$ is solvable by radicals over $\mathbb{Q}$.

Theorem 4.12. For each positive $n \in \mathbb{Z}$, $x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1$ is solvable by radicals over $\mathbb{Q}$.

Theorem 4.13. Every quadratic polynomial $p(x) \in \mathbb{Q}[x]$ is solvable by radicals over $\mathbb{Q}$.

Problem 4.14. Let $p(x) = x^6 - 3x^3 - 1$. Show that $p(x)$ is solvable by radicals over $\mathbb{Q}$.

Chapter 5

Rings

Our overarching goal, laid out in Chapter 2, is to show that there are quintic polynomials whose roots are *not* expressible in terms of its coefficients using just the operations of addition, subtraction, multiplication, division, and the extraction of roots (thus implying that there is no "quintic formula" that is analogous to the quadratic formula). We decided that we would say that such polynomials are *not* solvable by radicals, and in Chapter 4, we finally were able to write down a formal definition of this term. We also proved there are many polynomials that are solvable by radicals. But, how do we show that a polynomial is *not* solvable by radicals? We start by taking a closer look at polynomials.

5.1 Abstract rings

As we investigate polynomials, it will be useful to harness (and abstract) the algebraic properties that they possesses. For example, if we add two polynomials in $\mathbb{Q}[x]$, we obtain a polynomial that is again in $\mathbb{Q}[x]$, and similarly for multiplication. Let's explore the structure of $F[x]$ in general (where F is any field).

Definition 5.1. Let F be a field. The structure $(F[x], +, \cdot)$ consists of the set $F[x]$ together with the operations $+$ and $\cdot$ defined as follows. Let $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ and $q(x) = b_0 + b_1 x + \cdots + b_n x^n$ with $m \leq n$.

- **Addition:** $p(x)+q(x) = (a_0+b_0)+(a_1+b_1)x+\cdots+(a_n+b_n)x^n$ (where $a_k = 0$ when $k > m$).
- **Multiplication:** $p(x) \cdot q(x) = \displaystyle\sum_{k=0}^{m+n} c_k x^k$ where $c_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_{k-1} b_1 + a_k b_0$.

We often refer to the entire structure $(F[x], +, \cdot)$ as simply $F[x]$.

In the definition of polynomial multiplication above, $p(x) \cdot q(x)$ is just the result of applying the distributive law repeatedly to $(a_0 + a_1 x + \cdots + a_m x^m) \cdot (b_0 + b_1 x + \cdots + b_n x^n)$ and then grouping according to the powers of x. We will see that the operations of polynomial addition and multiplication have many familiar properties; let's prove a couple.

Problem 5.2. Using the definitions of polynomial addition and multiplication together with properties of fields, prove that for all fields F, both addition and multiplication in $F[x]$ are commutative.

Problem 5.3. Let F be any field. Find an additive identity for $F[x]$, and prove that it works. Also, if $p(x) = a_0 + a_1x + \cdots + a_mx^m$ is an arbitrary polynomial in $F[x]$, find its additive inverse, and prove that it works.

Problem 5.4. Which elements of $\mathbb{Q}[x]$ have a multiplicative inverse? Which do not? Justify your answer.

As should be becoming clear, many of the properties of F transfer to $F[x]$ (but not all!). Let's record some of those properties in a fact, which we will not prove. The existence of multiplicative inverses is notably absent.

Fact 5.5. Let F be any field. The following are true for $F[x]$.

- **Addition Laws:** Addition is associative and commutative. There is a unique additive identity, namely the constant zero polynomial, and every polynomial $p(x)$ has a unique additive inverse, denoted $-p(x)$.
- **Multiplication Laws:** Multiplication is associative and commutative. There is a unique multiplicative identity, namely the constant polynomial 1.
- **Distributivity Laws:** For all $p(x), q(x), r(x) \in F[x]$, $p(x)(q(x)+r(x)) = p(x)q(x)+p(x)r(x)$ and $(q(x) + r(x))p(x) = q(x)p(x) + r(x)p(x)$.

5.1.1 Definition and first examples

Since $F[x]$ lacks multiplicative inverses for many of its elements, it does not form a field. Nevertheless, motivated by our desire to study polynomials, we will abstract the structure that is present so that we can prove theorems about polynomials over any field, instead of working one field at a time. However, before we do, it's worth noting that there are many other structures that are not fields but do satisfy the laws in Fact 5.5—perhaps the most prominent one is the integers $\mathbb{Z}$. We arrive at the definition of a ring.

Definition 5.6. A **ring** is a structure $(R, +, \cdot)$ consisting of a set R together with two binary operations $+$ and $\cdot$ (which we call *addition* and *multiplication*) such that for some element $0 \in R$ the following axioms hold.

- **Addition Axioms:** Addition is associative and commutative; the element 0 is an additive identity; every $x \in R$ has an additive inverse with respect to 0, denoted $-x$.
- **Multiplication Axioms:** Multiplication is associative.
- **Distributivity Axioms:** For all $x, y, z \in R$, $x(y + z) = xy + xz$ and $(y + z)x = yx + zx$.

In the case that multiplication is commutative, R is called a **commutative ring**, and in the case that there is a multiplicative identity, R is called a **ring with unity** (or ring with 1).

The notion of a ring is quite general, and the terminology "commutative ring" and "ring with unity" highlight some of the additional properties that $F[x]$ has, but arbitrary rings may not. But notice that fields have all of these properties *and more*. The next definition is meant to highlight this.

Definition 5.7. A **division ring** is a ring with unity such that every nonzero element has a multiplicative inverse.

Problem 5.8. Fill in each box of the table below with **Y**es or **N**o. Assume that + and · are defined "as usual" for each set.*

	ring	commutative ring	ring with unity	division ring	field
$\mathbb{Z}$					
$2\mathbb{Z}$					
$\mathbb{N}$					
$\mathbb{Q}$					
$\mathbb{H}$					
$\mathbb{Z}_6$					
$\mathbb{R}[x]$					
$\{a+bi \mid a,b \in \mathbb{Q}\}$					
$\{a+bi \mid a,b \in \mathbb{Z}\}$					

5.1.2 Basic properties

Many of the basic properties of fields hold also for rings, with essentially the same proofs, so we will just take them as fact.

Fact 5.9 (Compare with Fact 3.49)**.** Let R be a ring.

(1) The additive identity is unique. If there exists a multiplicative identity, it is unique.

(2) Additive inverses are unique. If an element has a multiplicative inverse, it is unique.

Fact 5.10 (Compare with Fact 3.50)**.** Let R be a ring.

(1) For all $x \in R$, $x \cdot 0 = 0 = 0 \cdot x$.

(2) For all $x, y \in R$, $(-x)y = -(xy)$ and $x(-y) = -(xy)$.

(3) If R contains at least two elements and has a multiplicative identity, then the additive and multiplicative identities are different, i.e. $0 \neq 1$.

*$2\mathbb{Z}$ denotes the even integers. The operations are usual integer addition and multiplication.

Let's explore one further property that fields possess but is not listed above: for all x and y in a field, if $xy = 0$, then $x = 0$ or $y = 0$.

Definition 5.11. Let R be a ring. An element $a \in R$ is called a **zero divisor** if a is nonzero and there exists a nonzero $b \in R$ such that $ab = 0$. A ring is called an **integral domain** if it is a commutative ring with unity containing at least two elements but *no zero divisors*.

As remarked above, fields do not have zero divisors, so every field is indeed an integral domain. However, the prototypical integral domain (which explains the choice of name) is $\mathbb{Z}$. Let's look for others.

Problem 5.12. For each of the following rings, determine if there are zero divisors, and if so, find them all. Is the ring an integral domain?

(1) $\mathbb{Z}_5$

(2) $\mathbb{Z}_{10}$

(3) $\mathbb{H}$

(4) $\mathbb{R}[x]$

When working with integral domains, the following property is key.

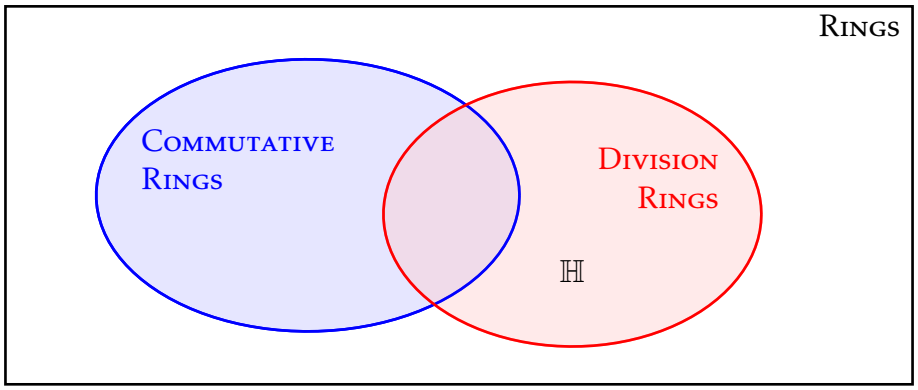
Theorem 5.13 (Cancellation Property)**.** Let R be an integral domain. For all $a, b, c \in R$, if $ab = ac$, then either $a = 0$ or $b = c$.

Problem 5.14. What properties of integral domains did you use in your proof of Theorem 5.13? Can you rewrite the theorem to be more general? Try.

Let's pause to collect and organize all of our new definitions.

Problem 5.15. Complete the following Venn Diagram by adding in shapes for each of the following terms. Try to provide examples that live in each of the gaps, but we have not encountered enough examples (in these notes) to cover all gaps yet.

- Fields
- Rings
- Commutative Rings
- Rings with unity
- Division Rings
- Integral domains



5.1.3 Units

Unless R is actually a division ring, not all elements of R will have a multiplicative inverse. Let's explore those elements that *do* have an inverse.

Definition 5.16. Let R be a ring with unity containing at least two elements. Then, $u \in R$ is called a **unit** if u has a multiplicative inverse. The set of all units in R is denoted $U(R)$.

Problem 5.17. For each of the following rings, find all of the units, i.e. determine $U(R)$.

(1) $\mathbb{Z}$

(2) $\mathbb{Z}_5$

(3) $\mathbb{R}$

(4) $\mathbb{R}[x]$

Problem 5.18. Consider the ring $\mathbb{Z}_{20}$. Find all units of $\mathbb{Z}_{20}$ and also find all zero divisors. What do you notice?

Problem 5.19. Let n be a positive integer. Make a conjecture about $U(\mathbb{Z}_n)$ by filling in the blank: $U(\mathbb{Z}_n) = \{a \in \mathbb{Z}_n \mid$ (fill in the blank) $\}$. What evidence do you have?

Theorem 5.20. Let R be a ring with unity containing at least two elements. If $u \in R$ is a unit, then u is *not* a zero divisor.

Problem 5.21. Either prove or disprove the *converse* of Theorem 5.20.

Theorem 5.22. Let R be a ring with unity containing at least two elements. Then $(U(R), \cdot)$ is a group.

5.2 An aside: matrix rings

Matrix rings are really the prototypical ring with unity. Although you may have only seen matrices with real entries, it turns out that we can do matrix arithmetic with other types of entries, e.g. entries from $\mathbb{C}$ or $\mathbb{Z}$. In fact, the usual matrix addition and multiplication makes sense when the entries come from any ring.

Definition 5.23. Let R be a ring and n a positive integer. Then $M_n(R)$ denotes the set of all $n \times n$ matrices whose entries come from R. The structure $(M_n(R), +, \cdot)$ consists of the set $M_n(R)$ of all $n \times n$ matrices whose entries come from R, together with the operations of usual matrix addition and matrix multiplication.

Problem 5.24. Provide examples of matrices satisfying each of the following conditions.

(1) $A \in M_3(\mathbb{C})$ but $A \notin M_3(\mathbb{R})$

(2) $B \in M_2(\mathbb{H})$ but $B \notin M_2(\mathbb{C})$

(3) $C \in M_2\left(\mathbb{Q}\left(\sqrt{5}\right)\right)$ but $C \notin M_2(\mathbb{Q})$

(4) $D \in M_2(\mathbb{R}[x])$ but $D \notin M_2(\mathbb{R})$

Problem 5.25. Verify that $M_2(\mathbb{Z})$ is closed under matrix multiplication.

The next fact shows that $M_n(R)$ is a ring with unity (for each positive n). Afterward, we will explore some of the other ring properties we discussed above.

Fact 5.26. Let R be any ring. The following are true for $M_n(R)$.

- **Addition Laws:** Addition is associative and commutative. There is a unique additive identity, namely the matrix with all entries equal to 0, and every matrix A has a unique additive inverse, denoted $-A$.
- **Multiplication Laws:** Multiplication is associative. There is a unique multiplicative identity, namely the matrix with 1's on the main diagonal and 0's everywhere else.
- **Distributivity Laws:** For all $A, B, C \in M_n(R)$, $A(B+C) = AB + AC$ and $(B+C)A = BA + CA$.

Problem 5.27. Is $M_2(\mathbb{R})$ commutative? Prove your answer.

Problem 5.28. Does $M_2(\mathbb{R})$ have zero divisors? Prove your answer.

The collection of units in a matrix ring forms a group with respect to matrix multiplication by Theorem 5.22. It is a very important object and even has a special name.

Definition 5.29. Let R be a ring and n a positive integer. The **general linear group** over the ring R, denoted $\mathrm{GL}_n(R)$, is the group of units in the ring $M_n(R)$.

Problem 5.30. Show that $\begin{bmatrix} i & 3 \\ 0 & i \end{bmatrix} \in \mathrm{GL}_2(\mathbb{C})$ by finding a multiplicative inverse for it. Also, find two different matrices in $M_2(\mathbb{C})$ that are *not* in $\mathrm{GL}_2(\mathbb{C})$.

5.3 Polynomial rings

Our study of rings was motivated by our desire to learn more about polynomials, and we now dive a little deeper into the theory of polynomial rings. Ultimately, we will focus on polynomials rings $F[x]$ where F is a field. In this section, we will see that $F[x]$ behaves in many ways like the integers $\mathbb{Z}$: $F[x]$ is an integral domain, there is a division algorithm for $F[x]$, there exists a greatest common divisor for polynomials, and there is a notion of primes and prime factorizations. Let's start with some important terminology.

Definition 5.31. Let R be a ring, and let $p(x) \in R[x]$ be a *nonzero* polynomial. If $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$, then n is called the **degree** of $p(x)$, denoted $\deg p(x)$. In words, $\deg p(x)$ is the highest power of x in $p(x)$ with a nonzero coefficient. The degree of the zero polynomial is undefined.

Problem 5.32. Determine the degree of each of the following polynomials.

(1) $q(x) = 4x^5 + 2x^2 + 5 - 8x^2 + 2x^5$ in the ring $\mathbb{Z}[x]$

(2) $r(x) = 4x^5 + 2x^2 + 5 - 14x^2 + 2x^5$ in the ring $\mathbb{Z}_6[x]$

(3) $p(t) = (3t^2 - \sqrt{2})(-1 + 2t - t^3)$ in the ring $\mathbb{R}[t]$

(4) $s(x) = (5-i)^8 - (5-s)^8$ in the ring $\mathbb{C}[s]$

The degree function is incredibly useful when working with polynomials—let's prove a couple of properties about it.

Theorem 5.33. Let R be a ring. If $p(x)$ and $q(x)$ are nonzero polynomials $R[x]$, then $\deg(p(x)+q(x)) \leq \max(\deg p(x), \deg q(x))$ or $\deg(p(x)+q(x))$ is undefined.

Problem 5.34. Give an example of polynomials $p(x), q(x) \in \mathbb{Q}[x]$ such that $\deg(p(x)+q(x)) < \max(\deg p(x), \deg q(x))$.

Theorem 5.35. Let D be an integral domain. If $p(x)$ and $q(x)$ are nonzero polynomials $D[x]$, then $\deg(p(x)q(x)) = \deg p(x) + \deg q(x)$, and in particular, $\deg(p(x)q(x))$ is defined.

Problem 5.36. Give an example of nonzero polynomials $p(x), q(x) \in \mathbb{Z}_{10}[x]$ such that $\deg(p(x)q(x)) \neq \deg p(x) + \deg q(x)$. Why does this not contradict Theorem 5.35?

Corollary 5.37. If D is an integral domain, then $D[x]$ is an integral domain.

5.3.1 Division algorithm

Here we explore what it means for one polynomial to divide another as well as the idea of a quotient and remainder for division. These should be familiar from previous classes for $\mathbb{R}[x]$, but here we see that they generalize to arbitrary $F[x]$ for F a field.

Definition 5.38. Let R be a ring, and let $a, b \in R$. We say that b **divides** a (or b is a **divisor** of a) if there exists some $q \in R$ such that $a = bq$.

Problem 5.39. Consider the polynomial $p(x) = x^2 - 1$ in $\mathbb{Q}[x]$.

(1) Does $x+1$ divide $p(x)$ in $\mathbb{Q}[x]$? Why or why not?

(2) Does $p(x)$ divide $x+1$ in $\mathbb{Q}[x]$? Why or why not?

(3) Does 3 divide $p(x)$ in $\mathbb{Q}[x]$? Why or why not?

Theorem 5.40. Let $p(x) \in R[x]$ with R a ring. If $c \in R$ and $(x-c)$ divides $p(x)$, then $p(c) = 0$.

Even if $b(x)$ does not divide $a(x)$, it can still be useful to perform the division to obtain a quotient and remainder.

Problem 5.41. Consider the polynomials $a(x) = x^4 + x^3 - 8x + 5$ and $b(x) = x^2 - 3$ in $\mathbb{Q}[x]$. Use polynomial long division to show that $b(x)$ does not divide $a(x)$. What is the quotient and what is the remainder? Write $a(x)$ as $a(x) = b(x)q(x) + r(x)$ for some $q(x), r(x) \in \mathbb{Q}[x]$ with $\deg r(x) < \deg b(x)$.

The "division algorithm" (Theorem 5.43) formalizes what results from long division. And, it turns out that it is true for polynomials over any field (not just $\mathbb{Q}$). The next lemma prepares for the proof of the division algorithm.

Lemma 5.42. Let F be a field, and let $a(x), b(x) \in F[x]$ with $\deg a(x) \geq \deg b(x)$. Assume that $a(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$ and $b(x) = b_0 + b_1 x + \cdots + b_m x^m$ with $b_m \neq 0$. Set $a_1(x) = a(x) - b(x) a_n b_m^{-1} x^{n-m}$. Then $\deg a_1(x) < \deg a(x)$, and $b(x)$ divides $a(x) - a_1(x)$.

Suppose we are trying to divide $b(x)$ into $a(x)$. How do we find the quotient and the remainder? Well, if the degree of $a(x)$ is smaller than the degree of $b(x)$ there is nothing to do (and if $a(x) = 0$, there is also nothing to do). Otherwise, we can use the previous lemma to produce a polynomial $a_1(x)$ such that $\deg a_1(x) < \deg a(x)$ and $b(x)$ divides $a(x) - a_1(x)$, or in other words, $a(x) - a_1(x) = b(x)q_1(x)$ for some $q_1(x)$. Now, suppose we repeat the process and divide $b(x)$ into the resulting $a_1(x)$ to produce $a_2(x)$ and $q_2(x)$. Continuing in this fashion, we produce $a_2, q_2, a_3, q_3 \ldots, a_k, q_k$, stopping once the degree of $a_k(x)$ becomes smaller than the degree of $b(x)$ (or $a_k(x) = 0$). In total, we get something like the following.

$$\begin{aligned} a(x) - a_1(x) &= b(x)q_1(x) \\ a_1(x) - a_2(x) &= b(x)q_2(x) \\ a_2(x) - a_3(x) &= b(x)q_3(x) \\ &\vdots \\ a_{k-1}(x) - a_k(x) &= b(x)q_k(x) \end{aligned}$$

Adding the above equations together and moving things around, we arrive at

$$a(x) = b(x)\left(q_1(x) + q_2(x) + q_3(x) + \cdots + q_k(x)\right) + a_k(x)$$

with the degree of $a_k(x)$ being less than the degree of $b(x)$. Thus, $a_k(x)$ is the remainder and $q_1(x) + \cdots + q_k(x)$ the quotient. This is the rough idea behind the division algorithm.

Theorem 5.43 (Division algorithm for $F[x]$)**.** Let F be a field, and let $a(x), b(x) \in F[x]$ with $b(x) \neq 0$. Then there exist $q(x), r(x) \in F[x]$ such that

$$a(x) = b(x)q(x) + r(x)$$

with $\deg r(x) < \deg b(x)$ or $r(x) = 0$.

The division algorithm is the theoretical analogue of long division. If you want to divide concrete polynomials, use long division, but if you want to prove something about divisibility for arbitrary polynomials, use the division algorithm. It is often used to prove a polynomial $b(x)$ actually divides another polynomial $a(x)$. The strategy is to apply the division algorithm to produce the equation $a(x) = b(x)q(x) + r(x)$ (with $\deg r(x) < \deg b(x)$ or $r(x) = 0$) and then use this to show that, in fact, $r(x) = 0$, implying that $a(x) = b(x)q(x)$ as desired. Let's try using this approach to prove the converse of Theorem 5.40.

Theorem 5.44. Let $p(x) \in F[x]$ for F a field. If $c \in F$ and $p(c) = 0$, then $(x - c)$ divides $p(x)$.

Problem 5.45. Consider the polynomial $p(x) = x^2 + x + 3$ in $\mathbb{Z}_5[x]$. Compute $p(c)$ for each $c \in \mathbb{Z}_5$, and use the results to determine which polynomials of the form $(x - c)$ divide $p(x)$. Factor $p(x)$ into a product of degree 1 polynomials in $\mathbb{Z}_5[x]$, if possible.

Problem 5.46. Consider the polynomial $p(x) = x^2 + x + 1$ in $\mathbb{Z}_5[x]$. Explain why $p(x)$ cannot be factored into a product of degree 1 polynomials in $\mathbb{Z}_5[x]$.

5.3.2 Greatest common divisors

The fact that there is a division algorithm for $F[x]$ (Theorem 5.43) is a rather special property for a ring to possess, and it has several important consequences. The first one we'll explore is the existence of a "greatest common divisor" for two polynomials, and our first order of business is to try to decide on a reasonable definition of this.

Problem 5.47. What are the common divisors of 6 and -9 in $\mathbb{Z}$? Which one is the greatest common divisor?

Problem 5.48. Consider the polynomials $a(x) = 2x^2 - 2$ and $b(x) = 2x^2 + 2x - 4$ in $\mathbb{Q}[x]$.

(1) Show that $x-1$ is a common divisor of $a(x)$ and $b(x)$ by finding $q(x), s(x) \in \mathbb{Q}[x]$ such that $a(x) = (x-1)q(x)$ and $b(x) = (x-1)s(x)$.

(2) Show that $-2(x-1)$ is a common divisor of $a(x)$ and $b(x)$ by finding $q(x), s(x) \in \mathbb{Q}[x]$ such that $a(x) = -2(x-1)q(x)$ and $b(x) = -2(x-1)s(x)$.

(3) Show that $100(x-1)$ is a common divisor of $a(x)$ and $b(x)$ by finding $q(x), s(x) \in \mathbb{Q}[x]$ such that $a(x) = 100(x-1)q(x)$ and $b(x) = 100(x-1)s(x)$.

Which one, if any, would be a good choice as the "greatest common divisor"?

The previous problem highlights that there are several (actually, infinitely many) choices for the "greatest common divisor" of two polynomials. Our choice for which one we call the greatest common divisor is, in some sense, the simplest one.

Definition 5.49. A polynomial $p(x)$ of degree n is called **monic** if the coefficient of x^n (i.e. the leading coefficient) is 1.

For example, $7 - 2x + x^2$ is monic, since the coefficient of x^2 is 1. However, neither $7 - 2x + 3x^2$ nor $7 - 2x - x^2$ are monic.

Definition 5.50. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. A polynomial $d(x) \in F[x]$ is called a **greatest common divisor** of $a(x)$ and $b(x)$ if

(1) $d(x)$ is monic,

(2) $d(x)$ divides both $a(x)$ and $b(x)$,

(3) if $h(x)$ divides both $a(x)$ and $b(x)$, then $h(x)$ divides $d(x)$.

Thus, in Problem 5.48, the greatest common divisor of the polynomials $a(x)$ and $b(x)$ is $x-1$. That said, we don't yet know that a greatest common divisor always exists, but let's start by showing that if one exists, there is only one.

Lemma 5.51. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. If $d_1(x)$ and $d_2(x)$ are greatest common divisors of $a(x)$ and $b(x)$, then $d_1(x) = d_2(x)$.

We now work towards the existence of a greatest common divisor for arbitrary polynomials in $F[x]$ (for arbitrary fields). The proof of this result is tightly tied to analyzing certain combinations of the polynomials $a(x)$ and $b(x)$. Let's explore this a bit.

Problem 5.52. Consider the polynomials $a(x) = 2x^2 - 2$ and $b(x) = 2x^2 + 2x - 4$ in $\mathbb{Q}[x]$. Define

$$I = \{f(x)a(x) + g(x)b(x) \mid f(x), g(x) \in \mathbb{Q}[x]\}.$$

(1) Write down 5 different polynomials that are in the set I.

(2) Show that $x - 1$ divides an arbitrary polynomial in I.

The idea behind the second part of Problem 5.52 can be used to prove the following general result about sets like I.

Theorem 5.53. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. Define

$$I = \{f(x)a(x) + g(x)b(x) \mid f(x), g(x) \in F[x]\}.$$

If $h(x)$ divides both $a(x)$ and $b(x)$, then $h(x)$ divides every $c(x) \in I$.

The existence and uniqueness of greatest common divisors in $F[x]$ is presented in the following fact.

Fact 5.54. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. There exists a unique greatest common divisor of $a(x)$ and $b(x)$, and if $d(x)$ is the greatest common divisor, then

$$d(x) = f(x)a(x) + g(x)b(x),$$

for some $f(x), g(x) \in F[x]$.

The proof of this fact is interesting, but let's content ourselves to just outline it. The approach is fairly straight forward. Define

$$I = \{f(x)a(x) + g(x)b(x) \mid f(x), g(x) \in F[x]\}.$$

Theorem 5.53 tells us that every common divisor of $a(x)$ and $b(x)$ is a divisor of every polynomial in I. Thus, if I contains a monic, common divisor of $a(x)$ and $b(x)$, it must be the greatest common divisor. So, we look for a common divisor of $a(x)$ and $b(x)$ in I. And to do this, the key idea is to choose a polynomial of smallest degree in I.

Let m be the smallest degree of all nonzero polynomials in I (which exists by the Well-Ordering Property of the natural numbers). Choosing any polynomial of degree m in I, we can divide out the leading coefficient to get a *monic* polynomial $d(x)$, which we can show is still in I. The polynomial $d(x)$ will be the greatest common divisor.

To see that $d(x)$ divides $a(x)$, we use Theorem 5.43 (the division algorithm) to write $a(x) = d(x)q(x) + r(x)$ for $q(x), r(x) \in F[x]$ with $\deg r(x) < \deg d(x)$ or $r(x) = 0$. Towards a contradiction, assume $r(x) \neq 0$. Now, since $d(x) \in I$, there exist $f_d(x), g_d(x) \in F[x]$ such that

$$\begin{aligned} r(x) &= a(x) - d(x)q(x) \\ &= a(x) - [f_d(x)a(x) + g_d(x)b(x)]q(x) \\ &= [1 - f_d(x)q(x)]a(x) + [-g_d(x)q(x)]b(x) \\ &\in I. \end{aligned}$$

Since $\deg r(x) < \deg d(x)$, this contradicts the fact that $d(x)$ had the smallest possible degree of all polynomials in I. Thus, $r(x) = 0$, and $d(x)$ divides $a(x)$. A similar argument shows that $d(x)$ also divides $b(x)$, so $d(x)$ is a monic, common divisor of $a(x)$ and $b(x)$. And, Theorem 5.53 shows that $d(x)$ is a greatest common divisor. But then, it is the unique greatest common divisor by Lemma 5.51.

Using Fact 5.54, we can rewrite the set I is a very nice way.

Corollary 5.55. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. Define

$$I = \{f(x)a(x) + g(x)b(x) \mid f(x), g(x) \in F[x]\}.$$

If $d(x)$ is the greatest common divisor of $a(x)$ and $b(x)$, then $I = \{p(x)d(x) \mid p(x) \in F[x]\}$.

With similar ideas as in the proof of Fact 5.54, one can prove the following fact that characterizes the greatest common divisor in several different ways.

Fact 5.56. Let F be a field, and let $a(x), b(x) \in F[x]$ be nonzero polynomials. Define

$$I = \{f(x)a(x) + g(x)b(x) \mid f(x), g(x) \in F[x]\}.$$

For any polynomial $d(x) \in F[x]$, the following are equivalent:

(1) $d(x)$ is the greatest common divisor of $a(x)$ and $b(x)$;

(2) $d(x)$ is a monic common divisor of $a(x)$ and $b(x)$, and $d(x) \in I$;

(3) $d(x)$ is a monic, and $I = \{p(x)d(x) \mid p(x) \in F[x]\}$.

So, you may be wondering: how do we compute the greatest common divisor of two polynomials? First think about how you would compute the greatest common divisor of 168 and 180. Really, think about it... Many people will factor both 168 and 180 into primes and then multiply the prime factors they have in common. This works for integers, and in fact, it will also work for polynomials once we develop the notion of a "prime polynomial." However, there is another approach, which in general is way more efficient: the Euclidean Algorithm. We will not develop it here, but you are encouraged to look it up (perhaps starting on Wikipedia).

5.3.3 Irreducible polynomials

We now develop the analogous notion of a prime number for polynomials, which will be an *irreducible* polynomial. The concept of irreducibility makes sense quite generally, so we start by defining it for any integral domain. Recall that, by Corollary 5.37, $F[x]$ is always an integral domain when F is a field.

To motivate the definition, think about how prime integers are defined: $p \in \mathbb{Z}$ is prime if (1) $p > 1$ and (2) $p = ab$ implies that $a = \pm 1$ or $b = \pm 1$. Since the units of $\mathbb{Z}$ are precisely ± 1, the second condition could be rewritten as "$p = ab$ implies that a or b is a unit." Also, since we don't want 1 (or -1) to be considered prime, the first condition is mostly captured by ensuring that "p is not zero and not a unit."

Definition 5.57. Let D be an integral domain. An element $p \in D$ is **irreducible** if

- $p \neq 0$ and $p \notin U(D)$, and
- for all $a, b \in D$, $p = ab$ implies $a \in U(D)$ or $b \in U(D)$.

The element p is **reducible** if it is not irreducible; that is if $p = 0$, $p \in U(D)$, or there exist $a, b \in D$ such that $p = ab$ and $a, b \notin U(D)$.

Problem 5.58. Use Definition 5.57 to show that a field has no irreducible elements.

Problem 5.59. What are the irreducible elements in $\mathbb{Z}$?

In order to investigate irreducibility in an integral domain D, we need to know its units. Our overarching goal is to better understand polynomials, so let's start there.

Theorem 5.60. Let F be a field. Then $p(x)$ is a unit in $F[x]$ if and only if $\deg p(x) = 0$.

Let's rewrite our definition of reducibility in a more useable form for polynomials.

Theorem 5.61. Let F be a field, and let $p(x)$ be a nonconstant polynomial in $F[x]$. Then $p(x)$ is reducible if and only if $\deg p(x) > 0$ and there exist polynomials $a(x), b(x) \in F[x]$ such that $p(x) = a(x)b(x)$ with $\deg a(x) < \deg p(x)$ and $\deg b(x) < \deg p(x)$.

Problem 5.62. Determine if $p(x)$ is reducible or irreducible in the given ring. If it's reducible, write down a factorization.

(1) $p(x) = x^2 + 1$ in $\mathbb{C}[x]$

(2) $p(x) = x^2 + 1$ in $\mathbb{Q}[x]$

(3) $p(x) = x^2 + 1$ in $\mathbb{Z}_2[x]$

(4) $p(x) = x^2 + 1$ in $\mathbb{Z}_3[x]$

Let's catalog a couple of general irreducibility/reducibility results for polynomials of small degree.

Theorem 5.63. Let F be a field. If $\deg p(x) = 1$, then $p(x)$ is irreducible.

Theorem 5.64. Let F be a field. If $\deg p(x) = 2, 3$, then $p(x)$ is reducible if and only if $p(x)$ has a root in F.

Problem 5.65. Determine if $p(x)$ is reducible or irreducible in the given ring. If it's reducible, write down a factorization.

(1) $p(x) = x^3 - 2$ in $\mathbb{Q}[x]$

(2) $p(x) = x^3 - 2$ in $\mathbb{Z}_5[x]$

Problem 5.66. Determine if each of the following polynomials are reducible or irreducible in the given ring.

(1) $p(x) = x^3 - 8$ in $\mathbb{Q}[x]$

(2) $p(x) = x^3 - 8$ in $\mathbb{Z}_5[x]$

(3) $r(x) = x^4 - 8x^2 + 15$ in $\mathbb{Q}[x]$

(4) $r(x) = x^4 - 8x^2 + 15$ in $\mathbb{Z}_5[x]$

To solidify the analogy between irreducible elements and primes, let's prove a factorization theorem.

Theorem 5.67. If F is a field, then any polynomial of positive degree in $F[x]$ can be written as a product of polynomials that are irreducible in $F[x]$.

As you know, in the integers every number greater than or equal to 2 can be factored into a product of primes in a way that is *unique up to reordering the factors*. There is a similar uniqueness result for polynomials: any polynomial of positive degree in $F[x]$ can be written as a product of irreducible polynomials in a way that is unique up to reordering the factors and multiplying each factor by a unit.

Problem 5.68. Let $p(x) = 6x^4 - 7x^3 + 15x^2 - 21x - 9$. Then the following are two different factorizations of $p(x)$ into irreducibles in $\mathbb{Q}[x]$:

- $p(x) = (2x-3)(3x+1)(x^2+3)$, and
- $p(x) = (x+\frac{1}{3})(2x^2+6)(3x-\frac{9}{2})$.

Explain why the factorizations are the same after possibly reordering the factors and multiplying each factor by a unit.

5.4 Subrings

We now return to general ring theory. As with groups and fields, the notion of a subring is fundamental.

Definition 5.69. Let $(R,+,\cdot)$ be a ring, and let S be a subset of R. Then S is a **subring** of R if S is a ring in its own right with respect to operations $+$ and $\cdot$ *inherited from* R.

As with fields, many of the properties of the ring R automatically pass to a subset S (e.g. associativity), leaving only a handful of the ring axioms to actually be verified.

Theorem 5.70. Let R be a ring, and let $S \subseteq R$. Then S is a subring of R if and only if

(1) S is nonempty;

(2) for all $x, y \in S$, $x+y \in S$ and $xy \in S$; and

(3) for all $x \in S$, $-x \in S$.

Problem 5.71. Determine if each of the following subsets of $\mathbb{Q}[x]$ are actually subrings.

(1) $A = \{p(x) \mid p(x) = c \text{ for some } c \in \mathbb{Q}\}$ (i.e. the set of constant polynomials)

(2) $B = \{p(x) \mid p(x) = 0 \text{ or } \deg p(x) \leq 1\}$ (i.e. the set of linear polynomials)

(3) $\mathbb{Z}[x]$

(4) $I = \{f(x)x^2 + g(x)(1+x^5) \mid f(x), g(x) \in \mathbb{Q}[x]\}$

Problem 5.72. Explain why $\mathbb{Z}_5$ is *not* a subring of $\mathbb{Z}$.

Examples of subrings of $\mathbb{C}$ include $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, and $\mathbb{Q}(i)$. These examples can, in turn, be used to create subrings of polynomial rings and matrix rings .

Theorem 5.73. If S is a subring of R, then

(1) $S[x]$ is a subring of $R[x]$, and

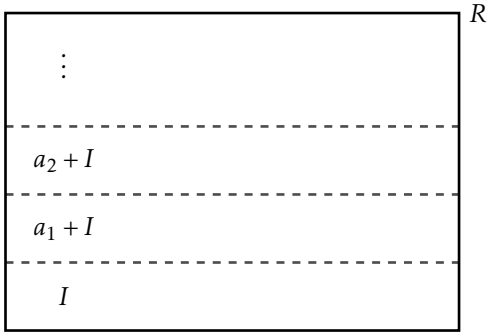
(2) $M_n(S)$ is a subring of $M_n(R)$.

5.5 Ideals and quotients

We now turn our attention to a special class of subrings known as *ideals*. The motivation for studying ideals of a ring is the same as for studying normal subgroups of a group: they give rise to quotients.

Let's explore the extra properties that a subring might need to ensure that the set of cosets can be given the structure of a ring. To start out, let's just assume that I is an *additive subgroup* of the ring R. Since $(R,+)$ is abelian, I is automatically a normal subgroup of $(R,+)$.

Now, let's consider the set of cosets of I in R, which we write as R/I. Recall that $R/I = \{a+I \mid a \in R\}$ where $a+I = \{a+y \mid y \in I\}$. Remember, that the set of cosets will partition R. One way to picture this is given below—it's followed by a couple of important properties about R/I from group theory.



Fact 5.74. Let I be an additive subgroup of a ring R. Then for all $a,b \in R$

(1) $a+I = b+I$ if and only if $a-b \in I$ if and only if $a \in b+I$; and

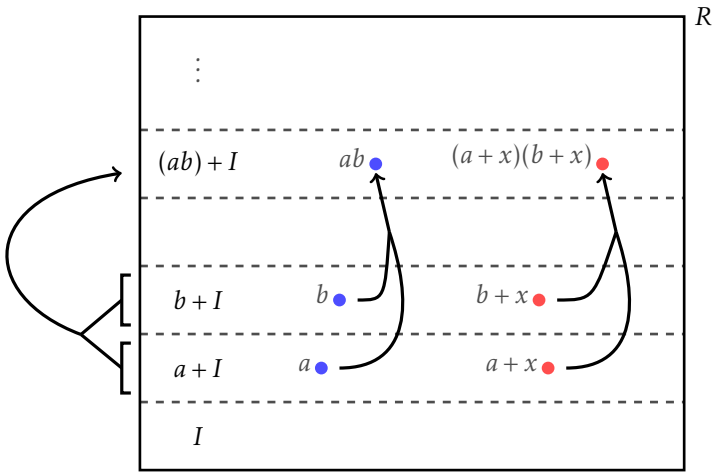
(2) either $(a+I)\cap(b+I) = \emptyset$ or $a+I = b+I$.

The goal is to understand when R/I can be given the structure of a ring. To do this, we need to decide how to add and multiply cosets. We would like to define $(a+I)+(b+I) = (a+b)+I$ and $(a+I)(b+I) = ab+I$, but the worry is that these operations are not well-defined. That is, the coset $a+I$ goes by many names (since $a+I = a'+I$ for every $a' \in a+I$), so we have to make sure that our definitions for the operations do not depend on which names we use for the cosets.

Now, as mentioned before, I is a *normal* subgroup of $(R,+)$, so we know that coset addition is well-defined. Let's see what we need for multiplication. Fix two arbitrary cosets $a+I$ and $b+I$. Then, for all $x,y \in I$, $a+I = (a+x)+I$ and $b+I = (b+y)+I$. Thus, in order for coset multiplication to be well-defined, we need to ensure that

$$ab+I = (a+x)(b+y)+I \text{ for all } a,b \in R \text{ and all } x,y \in I.$$

The desired picture is as follows.



After distributing, we have $ab+I = ab+ay+xb+xy+I$, which simplifies to $I = ay+xb+xy+I$. By Fact 5.74, we see that what we really need to ensure is that

$$ay+xb+xy \in I \text{ for all } a,b \in R \text{ and all } x,y \in I.$$

In particular, this has to be true when $a = b = 0$, which implies that $xy \in I$ for all $x,y \in I$, so I needs to be closed under multiplication, hence a subring. So, let's assume that I is a subring. Then, since $xy \in I$, $ay+xb+xy \in I$ reduces to $ay+xb \in I$. So, assuming that I is a subring, our previous condition becomes

$$ay+xb \in I \text{ for all } a,b \in R \text{ and all } x,y \in I.$$

Now, if a is arbitrary and $b = 0$, then we see that $ay \in I$ for all $a \in R$ and $y \in I$. Similarly, we find that $xb \in I$ for all $b \in R$ and $x \in I$. These are new properties. In words, I must be closed under (left and right) multiplication by elements from R.

In conclusion, if I is a subring that is closed under multiplication by elements from R then the above addition and multiplication for R/I is well-defined, making R/I a ring. The converse is also true. As such, we give these special subrings a special name.

Definition 5.75. Let R be a ring, and let $I \subseteq R$. Then I is an **ideal** of R if

(1) I is a subring; and

(2) for all $r \in R$ and all $a \in I$, both $ra \in I$ and $ar \in I$.

Let's also summarize our work above about defining operations on the quotient R/I.

Fact 5.76. Let R be a ring and let I be an ideal of R. Then R/I is a ring under the binary operations defined as follows. For all $a, b \in R$,

- $(a+I)+(b+I) = (a+b)+I$;
- $(a+I)(b+I) = (ab)+I$.

Problem 5.77. For each subset of the given ring, determine if the subset is an ideal, a subring, or neither.

	ideal	subring	neither
$\mathbb{Z} \subset \mathbb{Q}$			
$2\mathbb{Z} \subset \mathbb{Z}$			
$\{\text{odd integers}\} \subset \mathbb{Z}$			
$\{0,3,6,9\} \subset \mathbb{Z}_{12}$			
$\{p(x) \mid p(0) = 0\} \subset \mathbb{Q}[x]$			
$\{\text{constant polynomials}\} \subset \mathbb{Q}[x]$			

Problem 5.78. Let $I = \{(x^2+1)p(x) \mid p(x) \in \mathbb{Q}[x]\}$.

(1) Show that I is an ideal of $\mathbb{Q}[x]$.

(2) Write out 5 elements of I, each with a different degree.

(3) Explain why I contains polynomials of every degree larger than or equal to 2.

(4) Let $a(x) = x^4 + 3x + 5$. Write out 5 elements of the coset $a(x) + I$.

(5) Find some $b(x)$ in the coset $a(x) + I$ such that $\deg b(x) = 1$.

(6) Explain why $(x+I)^2 = -1 + I$ in the ring $\mathbb{Q}[x]/I$.

Problem 5.79. Let $I = \{(x^2+1)p(x) \mid p(x) \in \mathbb{Q}[x]\}$. Show that every coset $a(x) + I \in \mathbb{Q}[x]/I$ can be represented as $a(x) + I = r(x) + I$ for some $r(x) \in \mathbb{Q}[x]$ where $\deg r(x) < 2$ or $r(x) = 0$.

Problem 5.80. Recall that $6\mathbb{Z} = \{6z \mid z \in \mathbb{Z}\}$.

(1) Show that $6\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

(2) Show that $a + I = b + I$ if and only if $a \equiv_6 b$.

(3) Show that for all $a + I \in \mathbb{Z}/6\mathbb{Z}$, $a + I = r + I$ for some $r \in \mathbb{Z}$ with $0 \leq r < 6$.

Problem 5.81. Let $I = \{3q \mid q \in \mathbb{Q}\}$.

(1) Show that I is an ideal of $\mathbb{Q}$.

(2) Show that $1 \in I$, and use this to explain why $I = \mathbb{Q}$.

Let's record some observations from the previous problems.

Theorem 5.82. Let R be a commutative ring, and let $a \in R$. The set $I = \{ar \mid r \in R\}$ is an ideal of R.

In the previous theorem, the set $\{ar \mid r \in R\}$ should be thought of as the set of all multiples of a, and it is often denoted aR (as in $2\mathbb{Z}$).

Theorem 5.83. Assume R is a ring with unity. Let I be an ideal of R. If I contains a unit of R, then $I = R$.

Theorem 5.84. Let R be a ring. Then $\{0\}$ and R are ideals of R.

Theorem 5.85. Assume R is a commutative ring with unity and $1 \neq 0$. Then R is a field if and only if the only ideals of R are $\{0\}$ and R.

Theorem 5.86. Let I be an ideal of a ring R.

(1) If R is a commutative ring, then R/I is commutative ring.

(2) If R is a ring with unity, then R/I is a ring with unity.

5.5.1 Generating ideals

As with groups and fields, we will want to generate subobjects from subsets. Generating ideals will be more useful for us that subrings, so we will only focus on ideals. Regarding notation, it is common to use (A) for the ideal generated by A instead of $\langle A \rangle$, and we will follow that convention. We begin with intersections.

Theorem 5.87. If I and J are ideals of a ring R, then $I \cap J$ is an ideal of R.

This can be generalized to arbitrary intersections.

Theorem 5.88. If $\mathcal{C}$ is any collection of ideals of a ring R, then the intersection of all ideals from $\mathcal{C}$ is again an ideal of R.

Now, if A is any subset of R, we can let $\mathcal{C}$ be the collection of all ideals containing A, to see that the intersection of all ideals containing A is an ideal, and it must be the smallest ideal containing A. This leads to the following definition.

Definition 5.89. Suppose A is a subset of a ring R. The **ideal of R generated by** A, denoted (A), is defined to be the intersection of all ideals containing A. An ideal generated by one element is a **principal ideal**. If $A = \{a_1, \ldots a_k\}$, we often write $(a_1, \ldots a_k)$ in place of (A).

Problem 5.90. Consider the ring $\mathbb{Z}$.

(1) Recall that $2\mathbb{Z}$ is an ideal of $\mathbb{Z}$. Now use the definition of (2) to explain why $(2) \subseteq 2\mathbb{Z}$.

(2) Use that (2) is an ideal containing 2 to explain why $2\mathbb{Z} \subseteq (2)$. Conclude that $(2) = 2\mathbb{Z}$.

(3) Use that $(6, 10)$ is an ideal containing 6 and 10, to write down 5 elements of $(6, 10)$.

(4) Use the definition of $(6,10)$ to explain why $(6,10) \subseteq (2)$.

(5) Use Bézout's lemma (Fact 3.52) to show that $(2) \subseteq (6,10)$. Conclude that $(6,10) = (2)$.

(6) For arbitrary $m, n \in \mathbb{Z}$, do you think $(m,n) = (a)$ for some $a \in \mathbb{Z}$? Why or why not?

Let's work to abstract some of what we discovered in this problem.

Theorem 5.91. If R is a commutative ring with unity, then $(a) = \{ar \mid r \in R\}$.

Theorem 5.91 says that (a) is precisely the set of all multiples of a. Or, in other words, (a) is the set of all elements that are divisible by a. In particular, $(n) = n\mathbb{Z}$ in the ring $\mathbb{Z}$. But what about (m,n)?

Theorem 5.92. If $m, n \in \mathbb{Z}$ are nonzero, then $(m,n) = (d)$ where $d = \gcd(m,n)$.

In words, an ideal of $\mathbb{Z}$ that is generated by two elements can actually be generated by a single element. But more is true. The method for constructing the greatest common divisor for two elements can be easily adapted to show that *any* ideal of $\mathbb{Z}$ can be generated by a single element, which yields the following fact.

Fact 5.93. If I is any ideal of $\mathbb{Z}$, then I is a principle ideal. Moreover, if I is not the zero ideal, then $I = (d)$ if and only if d has the smallest possible absolute value among all nonzero elements of I.

In fact, a similar result holds in any ring with a division algorithm. Importantly for us, this applies to polynomial rings over fields. Let's prove the result for ideals generated by two elements and leave the general case as a fact.

Theorem 5.94. Let F be a field. If $a(x), b(x) \in F[x]$ are nonzero, then $(a(x), b(x)) = (d(x))$ where $d(x) = \gcd(a(x), b(x))$.

Fact 5.95. Let F be a field. If I is any ideal of $F[x]$, then I is a principle ideal. Moreover, if I is not the zero ideal, then $I = (d(x))$ if and only if $d(x)$ has the smallest possible degree among all nonzero elements of I.

Problem 5.96. Consider $a(x) = -x^2 - 3x + 10$, $b(x) = 2x^2 + 8x - 10$, and $c(x) = x^3 - 2$ in $\mathbb{Q}[x]$.

(1) Find a $d(x) \in \mathbb{Q}[x]$ such that $(a(x), b(x)) = (d(x))$. Is $(a(x), b(x)) = \mathbb{Q}[x]$? Explain.

(2) Find a $d'(x) \in \mathbb{Q}[x]$ such that $(a(x), c(x)) = (d'(x))$. Is $(a(x), c(x)) = \mathbb{Q}[x]$? Explain.

We saw that $\mathbb{Z}$ and $F[x]$ have a special property: every ideal is a principal ideal. This does not happen in every ring, as we'll see, so rings with this property get a special name.

Definition 5.97. An integral domain is called a **principal ideal domain (PID)** if every ideal is a principal ideal.

Thus, $\mathbb{Z}$ and $F[x]$ (for F a field) are examples of PIDs. Let's show that $\mathbb{Z}[x]$ is not.

Problem 5.98. Consider the set $I := \{2f(x) + xg(x) \mid f(x), g(x) \in \mathbb{Z}[x]\}$ in the ring $\mathbb{Z}[x]$. The set I is an ideal of $\mathbb{Z}[x]$; you do *not* need to prove this. Let's show that I is not principal. Towards a contradiction, assume that $I = (d(x))$ for some $d(x) \in \mathbb{Z}[x]$.

(1) Use the fact that $2 \in I = (d(x))$ to show that $d(x)$ is a constant polynomial and, moreover, that $d(x) = \pm 1, \pm 2$.

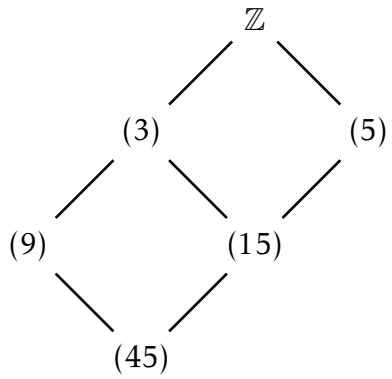
(2) Explain why every polynomial in I has a constant term that is even, and use this to show that in fact $d(x) = \pm 2$.

(3) Now, we also know that $x \in I = (d(x))$. Why is this a contradiction?

Let's return to $\mathbb{Z}$, and put together what we have learned about its ideals. First, by Fact 5.93, every ideal of $\mathbb{Z}$ is a principal ideal, so every ideal of $\mathbb{Z}$ is of the form (n) for some $n \in \mathbb{Z}$. Moreover, Theorem 5.91 tell us that (n) is just the set of all multiple of n, so $(n) = n\mathbb{Z}$. Thus, we know all of the ideals of $\mathbb{Z}$, and we know that they have a nice form. But how do they fit together? Let's explore this.

Theorem 5.99. Let $m, n \in \mathbb{Z}$. Then $(m) \subseteq (n)$ if and only if n divides m.

This theorem can be used to quickly draw portions of the lattice of ideals of $\mathbb{Z}$. For example, suppose we want to draw all of the ideals that contain the ideal (45). Every ideal is principal; suppose (n) is an ideal containing (45). By Theorem 5.99, n must divide 45. So, looking at all divisors of 45 (and noticing that $(-n) = (n)$), we get the following lattice.



Problem 5.100. Draw the the lattice of ideals of $\mathbb{Z}$ that contain the ideal (60).

Let's focus on the ideals of $\mathbb{Z}$ that are at the top of the lattice but below $\mathbb{Z}$.

Definition 5.101. An ideal M of a ring R is called a **maximal ideal** if $M \neq R$ and the only ideals containing M are M and R.

Theorem 5.102. An ideal I of $\mathbb{Z}$ is maximal if and only if $I = (p)$ for some prime $p \in \mathbb{Z}$.

5.6 Homomorphisms

When studying groups, we compared them using homomorphisms (and isomorphisms). We'll do the same for rings.

Definition 5.103. Let R and S be rings. A map $\phi : R \to S$ is called a **ring homomorphism** if the following are true for all $a, b \in R$:

(1) $\phi(a+b) = \phi(a) + \phi(b)$;

(2) $\phi(ab) = \phi(a)\phi(b)$.

If ϕ is a bijection, then ϕ is called an **isomorphism**, in which case, we say that R and S are **isomorphic rings** and write $R \cong S$.

Problem 5.104. Determine which of the following are ring homomorphisms. Explain.

(1) $\phi : \mathbb{Z} \to 3\mathbb{Z}$ defined by $\phi(n) = 3n$

(2) $\alpha : \mathbb{C} \to \mathbb{C}$ defined by $\alpha(a+bi) = a - bi$

(3) $\beta : \mathbb{Z} \to \mathbb{Z}$ defined by $\beta(a) = a^3$

(4) $f : \mathbb{Z}_3 \to \mathbb{Z}_3$ defined by $f(a) = a^3$

(5) $g : \mathbb{C} \to D_2(\mathbb{R})$ defined by $g(a+bi) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, where $D_2(\mathbb{R}) = \left\{ \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \mid x, y \in \mathbb{R} \right\}$

(6) $h : \mathbb{Q}[x] \to \mathbb{C}$ defined by $h(p(x)) = p(0)$

Problem 5.105. Recall that $\mathbb{Q}\left(\sqrt{5}\right)$ can be written as $\mathbb{Q}\left(\sqrt{5}\right) = \left\{a + b\sqrt{5} \mid a, b \in \mathbb{Q}\right\}$. Show that $\phi : \mathbb{Q}\left(\sqrt{5}\right) \to \mathbb{Q}\left(\sqrt{5}\right)$ defined by $\phi(a + b\sqrt{5}) = a - b\sqrt{5}$ is an isomorphism.

Definition 5.106. Let $\phi : R \to S$ be a ring homomorphism.

- The **kernel** of ϕ, denote $\ker(\phi)$, is defined to be $\ker(\phi) = \{a \in R \mid \phi(a) = 0\}$.
- The **image** of ϕ, denoted $\phi(R)$, is defined to be $\phi(R) = \{b \in S \mid b = \phi(a) \text{ for some } a \in R\}$.

Problem 5.107. Determine the kernel and image of each homomophism.

(1) $\alpha : \mathbb{Q}[x] \to \mathbb{C}$ defined by $\alpha(p(x)) = p(0)$

(2) $\beta : \mathbb{Z} \to \mathbb{Z}_5$ defined by $\beta(n) = n \pmod 5$

As with groups the kernel and image of a homomorphism has special properties.

Theorem 5.108. Let $\phi : R \to S$ be a ring homomorphism. Then $\ker(\phi)$ is an ideal of R, and $\phi(R)$ is a subring of S.

5.6.1 Isomorphism theorems

We have several isomorphism theorems for rings that mirror, exactly, what happed for groups. The proofs are essentially the same as before, so we will just state them as facts.

Fact 5.109 (First Isomorphism Theorem for Rings)**.** If $\phi : R \to S$ is a ring homomorphism, then $R/\ker(\phi) \cong \phi(R)$.

Problem 5.110. The map $\phi : \mathbb{Q}[x] \to \mathbb{C}$ defined by $\phi(p(x)) = p(i)$.

(1) Show that $\ker(\phi) = (x^2 + 1)$ (where $(x^2 + 1)$ is the ideal generated by $x^2 + 1$ in $\mathbb{Q}[x]$).

(2) Show that $\phi(R) = \mathbb{Q}(i)$.

(3) Conclude that $\mathbb{Q}[x]/(x^2 + 1) = \mathbb{Q}(i)$.

Appendix A

Hints

Below are some hints, which should be interpreted as possible (but not the only!) ways to get started.

Hint (Theorem 2.4)**.** You are solving $x^2 + bx + c = 0$. Try "completing the square" first; then solve for x.

Hint (Problem 3.9)**.** Multiplying a fraction by the complex conjugate of the denominator can be an effective way to simplify an expression.

Hint (Theorem 3.11)**.** Think back to changing from polar to rectangular coordinates (or parametrizing circles or solving triangles).

Hint (Theorem 3.12)**.** Try using Theorem 3.11 + trigonometric identities.

Hint (Problem 3.20)**.** You want to find a z such that $z^4 = \zeta_3$. You are working with powers (hence multiplication), so try writing z in the form $z = r\cos\theta + ir\sin\theta$. Now you can use Corollary 3.14 to simplify z^4 and compare with ζ_3. What can you deduce about r and θ?

Hint (Lemma 3.22)**.** Similar to Problem 3.20, try writing z in the form $z = r\cos\theta + ir\sin\theta$. Now, what does $z^n = 1$ imply about r and θ?

Hint (Lemma 3.23)**.** It may be helpful to draw some pictures first. Try plotting ζ_8, $(\zeta_8)^2$, $(\zeta_8)^3$, ..., $(\zeta_8)^8$, $(\zeta_8)^{14}$, $(\zeta_8)^{85}$. Now, you know by a previous problem that $(\zeta_n)^n = 1$, so also $(\zeta_n)^{2n} = 1$ and so on. Try (using the division algorithm) to write $k = qn + r$ for some $q, r \in \mathbb{Z}$ with $0 \le r \le n-1$ and plug that into $(\zeta_n)^k$.

Hint (Theorem 3.24)**.** You may want to view this as the following "if and only if" statement: z is an n^{th} root of $1 \iff z = (\zeta_n)^k$ for some $0 \le k \le n-1$. Now make use of the previous lemma and theorems you proved. Don't forget to explain why each of $1, \zeta_n, (\zeta_n)^2, \ldots, (\zeta_n)^{n-1}$ are all different.

Hint (Theorem 3.28)**.** Suppose that z is a root of $p(x)$. Then $p(z) = 0$, so $a_n z^n + a_{n-1} z^{n-1} + \cdots + a_2 z^2 + a_1 z + a_0 = 0$. This last equation is is just comparing two complex numbers—try taking the conjugate of both sides. Fact 3.5 is helpful.

Hint (Problem 3.40). You are trying to find $(a+b\sqrt{5})^{-1} = \frac{1}{a+b\sqrt{5}}$. Try multiplying top and bottom by the conjugate: $a-b\sqrt{5}$.

Hint (Theorem 3.50). For the first part, notice that $x \cdot 0 = x(0+0)$. For the last part, remember that the definition of a field ensures that F has at least two elements, so there is some $a \in F$ with $a \neq 0$. Now, what happens if $0 = 1$?

Hint (Theorem 3.53). The crux is to show that every nonzero element has a multiplicative inverse when n is prime. Let $a \in (\mathbb{Z}_n)^*$. You need to find some integer b such that $ab = 1$ modulo n. Now, since $a \in (\mathbb{Z}_n)^*$ and n is prime, $\gcd(a,n) = 1$. By Bézout's Lemma, there exist $k,l \in \mathbb{Z}$ such that $1 = ka + ln$. What happens when you consider the equation $1 = ka + lp$ modulo n?

Hint (Problem 3.57). If T_3 is a subfield, then, in particular, it is closed under multiplication, so it must be that $\alpha^2 \in T_3$. That means that $\alpha^2 = a + b\alpha$ for some $a, b \in \mathbb{Q}$. What does this imply?

Hint (Problem 3.64). Try following the approach in Example 3.60. First show $\{a+bi \mid a,b \in \mathbb{Q}\} \subseteq \mathbb{Q}(i)$ by showing that every subfield that contains $\mathbb{Q}$ and i must also contain $\{a+bi \mid a,b \in \mathbb{Q}\}$. To show the reverse containment, use the fact that $\{a+bi \mid a,b \in \mathbb{Q}\}$ is a subfield, by a previous problem.

Hint (Problem 3.67). Remember, in Problem 3.57(3), we saw that $\{a+b\alpha \mid a,b \in \mathbb{Q}\}$ is *not* a subfield of $\mathbb{C}$.

Hint (Problem 3.69). Use the previous theorem. To show $\mathbb{Q}\left(3-\sqrt{2}, 5+i\right) \subseteq \mathbb{Q}\left(\sqrt{2}, i\right)$, you need to show that $\mathbb{Q} \subset \mathbb{Q}\left(\sqrt{2}, i\right)$ and that $3-\sqrt{2}, 5+i \in \mathbb{Q}\left(\sqrt{2}, i\right)$. Then show the reverse containment in a similar way.

Hint (Theorem 4.12). Note that $x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1)$. Now use Theorem 3.24; note that $x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1$ should only have $n-1$ roots.

Hint (Problem 4.14). First find the roots of $z^2 - 3z - 1$. Then, for each of those roots, use Theorem 3.26 to solve for z. You should have 6 different roots in the end.

Hint (Theorem 5.20). Try a proof by contradiction. Assume that u is a unit and that u is a zero divisor. Now, what does the definition of being a zero divisor tell you about u?

Hint (Theorem 5.33). To get started, let $n = \deg p(x)$ and $m = \deg q(x)$, and then write $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$ and $q(x) = b_0 + b_1 x + \cdots + b_m x^m$ with $b_m \neq 0$. You want to understand the degree of $p(x) + q(x)$, so you need to determine the largest power of x in the sum $p(x) + q(x)$.

Hint (Theorem 5.35). As with the previous theorem, let $n = \deg p(x)$ and $m = \deg q(x)$, and then write $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$ and $q(x) = b_0 + b_1 x + \cdots + b_m x^m$ with $b_m \neq 0$. You need to determine the largest power of x in the product $p(x)q(x)$. What do you think is the largest power of x in the product $p(x)q(x)$? What is its coefficient, and how do you know it's not zero?

Hint (Corollary 5.37)**.** There are several things to verify to ensure that $D[x]$ is an integral domain, but we've talked about most of them already. The main thing that remains is to prove that $D[x]$ has no zero divisors—try a proof by contradiction. This is a corollary of Theorem 5.35, which means that it should be "not too hard" to prove using Theorem 5.35.

Hint (Theorem 5.43)**.** One approach is to polish up and fill in the gaps of the outline presented in the notes right before the statement of Theorem 5.43. A related, but slightly different, approach is to try using induction on the degree of $a(x)$.

Hint (Theorem 5.44)**.** Try using the division algorithm to write $a(x) = (x-c)q(x) + r(x)$ for some $q(x), r(x) \in F[x]$ with $\deg r(x) < \deg(x-c)$ or $r(x) = 0$. Now show that $r(x)$ must be the zero polynomial.

Hint (Lemma 5.51)**.** First, explain why $d_1(x)$ must divide $d_2(x)$ and why $d_2(x)$ must divide $d_1(x)$. Now return to the definition of "to divide" and see what you can write down.

Hint (Theorem 5.53)**.** Follow the definitions. Since $c(x) \in I$, it can be written a particular way. Then write down what it means for $h(x)$ to divide both $a(x)$ and $b(x)$. Combine.

Hint (Theorem 5.60)**.** For the forward direction, start with the definition of a unit and apply the degree function. For the reverse direction, what does $\deg p(x) = 0$ imply about $p(x)$? Can you explicitly write down the a multiplicative inverse for $p(x)$?

Hint (Theorem 5.64)**.** Consider using Theorem 5.40.

Hint (Theorem 5.67)**.** Consider using using strong induction on the degree of the polynomial. Let $\varphi(n)$ be the statement "every polynomial in $F[x]$ of degree n can be written as a product of polynomials that are irreducible in $F[x]$."

For the base case, you want to show that $\varphi(1)$ is true. Assume that $p(x) \in F[x]$ has degree 1. Then what?

Next, assume that $\varphi(k)$ is true for all $1 \leq k \leq n$. We need to show that $\varphi(n+1)$ is true. Assume that $p(x) \in F[x]$ has degree $n+1$. There are two cases to consider: $p(x)$ is irreducible or $p(x)$ is reducible. Keep going...

Hint (Problem 5.79)**.** Use the division algorithm to write $a(x) = (x^2+1)q(x) + r(x)$. What does this tell you?

Hint (Theorem 5.86)**.** Using Fact 5.76, you know that R/I is ring. So, for the first part, assume R is commutative, and use this to show R/I is commutative. The starting point is to choose two arbitrary elements of R/I, which would be something like $a+I$ and $b+I$ for $a, b \in R$. Now show that $(a+I)(b+I) = (b+I)(a+I)$ using the definition of multiplication in Fact 5.76.

Hint (Problem 5.80)**.** For the second part, remember that $a \equiv_6 b \iff a - b$ is a multiple of 6. For the last, use the division algorithm to write $a = 6q + r$. What does this imply?

Hint (Theorem 5.83)**.** By definition of an ideal, $I \subseteq R$, so what we really need to show is that $R \subseteq I$. Remember that I is closed under multiplication by elements of R. So, if $a \in I$, then $ra \in R$. Try to first show that $1 \in R$.

Hint (Theorem 5.85)**.** Theorem 5.83 should help with the forward direction. For the backward direction, let $a \in R^*$; you need to show a has an inverse. Try using Theorem 5.82: the set $I = \{ar \mid r \in R\}$ is an ideal. By assumption, $I = \{0\}$ or $I = R$. Which is it? Notice that if $I = R$, then $1 \in I$.

Hint (Problem 5.96)**.** Use Theorem 5.94. Theorem 5.83 may also be helpful.

Hint (Theorem 5.99)**.** Try using Theorem 5.91.