NOTES ON IDEALS

KEITH CONRAD

1. Introduction

Let R be a commutative ring (with identity). An *ideal* in R is an additive subgroup $I \subset R$ such that for all $x \in I$, $Rx \subset I$.

Example 1.1. For $a \in R$,

$$(a) := Ra = \{ra : r \in R\}$$

is an ideal. An ideal of the form (a) is called a *principal ideal* with *generator* a. We have $b \in (a)$ if and only if $a \mid b$. Note $(1) = R$.

An ideal containing an invertible element u also contains $u^{-1}u = 1$ and thus contains every $r \in R$ since $r = r \cdot 1$, so the ideal is R. This is why $(1) = R$ is called the *unit ideal*: it's the only ideal containing units (invertible elements).

Example 1.2. For a and $b \in R$,

$$(a, b) := Ra + Rb = \left\{ra + r'b : r, r' \in R\right\}$$

is an ideal. It is called the *ideal generated by* a and b. Note this is *not* $(a) \cup (b)$! That's like in group theory, where $\langle g, h\rangle$ consists of products of powers of g and h rather than being $\langle g\rangle \cup \langle h\rangle$.

More generally, for $a_1, \ldots, a_n \in R$ the set $(a_1, a_2, \ldots, a_n) = Ra_1 + \cdots + Ra_n$ is an ideal in R, called a *finitely generated ideal* or the ideal *generated by* $a_1, \ldots, a_n$. In some rings every ideal is principal, or more broadly every ideal is finitely generated, but there are also some "big" rings in which some ideal is not finitely generated.

It would be wrong to say an ideal is not principal if it is described with two generators: an ideal generated by several elements might be generated by fewer elements and even by one element (a principal ideal). For example, in $\mathbf{Z}$,

$$(6, 8) = 6\mathbf{Z} + 8\mathbf{Z} \stackrel{!}{=} 2\mathbf{Z}. \tag{1.1}$$

Both 8 and 6 are elements of the ideal $(6, 8)$, so $8 - 6 = 2$ is in the ideal. Hence every multiple of 2 is in the ideal, so $2\mathbf{Z} \subset (6, 8)$. Conversely, the ideal $(6, 8)$ is in $2\mathbf{Z}$ since every $6m + 8n$ is even. Thus $(6, 8) = (2)$ as ideals in $\mathbf{Z}$. This is analogous to group theory where a subgroup generated by two elements could be cyclic (having a single generator).

Remark 1.3. The elements of the ideal $(a, b) = aR + bR$ are all possible $ax + by$. This includes the multiples of a and the multiples of b, but (a, b) more than that in general: a typical element in (a, b) need not be a multiple of a or of b. Consider in $\mathbf{Z}$ the ideal $(6, 8) = 6\mathbf{Z} + 8\mathbf{Z} = 2\mathbf{Z}$: most even numbers are not multiples of 6 or of 8. Don't confuse the ideal (a, b) with the union $(a) \cup (b)$, which is usually *not* an ideal and in fact is not of much interest.

Generators of an ideal in a ring are analogous to a spanning set of a subspace of $\mathbf{R}^n$. But there is an important difference, illustrated by equation (1.1): all minimal spanning sets for a subspace of $\mathbf{R}^n$ have the same size (dimension of the subspace), but in $\mathbf{Z}$ the ideal of even numbers has minimal spanning sets $\{2\}$ and $\{6, 8\}$, which are of different sizes.

Example 1.4. For rings R and S, $R \times S$ is a ring with componentwise operations. The subsets $R \times \{0\} = \{(r, 0) : r \in R\}$ and $\{0\} \times S = \{(0, s) : s \in S\}$ are ideals in $R \times S$. Both are principal ideals: $R \times \{0\} = ((1, 0))$ and $\{0\} \times S = ((0, 1))$ in $R \times S$.

It is *not* obvious at first why the concept of an ideal is important. Here are three reasons why it is.

(1) Ideals in R are precisely the kernels of ring homomorphisms out of R, just as normal subgroups of a group G are precisely the kernels of group homomorphisms out of G. We will see why in Section 3.
(2) Ideals were first introduced, by Kummer, to restore unique factorization in certain rings where that property failed. He referred not to ideals as defined above but to "ideal numbers," somewhat in the spirit of the term "imaginary numbers."
(3) The study of commutative rings used to be called "ideal theory" (now it is called commutative algebra), so evidently ideals have to be a pretty central aspect of research into the structure of rings.

The following theorem says fields can be characterized by the types of ideals in it.

Theorem 1.5. *Let a commutative ring R not be the zero ring. Then R is a field if and only if its only ideals are (0) and (1).*

Proof. In a field, every nonzero element is invertible, so an ideal in the field other than (0) contains 1 and thus is (1). Conversely, if the only ideals are (0) and (1) then for all $a \neq 0$ in R we have $(a) = (1)$, and that implies $1 = ab$ for some b, so a has an inverse. Therefore all nonzero elements of R are invertible, so R is a field. □

2. Principal ideals

When is $(a) \subset (b)$? An ideal containing a also contains (a), and *vice versa*, so the condition $(a) \subset (b)$ is the same as $a \in (b)$, which is true if and only if $a = bc$ for some $c \in R$, which means $b \mid a$. Thus

$$(a) \subset (b) \iff b \mid a \text{ in } R.$$

Thus inclusion of one principal ideal in another corresponds to *reverse* divisibility of the generators, or equivalently divisibility of one number into another in R corresponds to *reverse* inclusion of the principal ideals they generate: $x \mid y$ in $R \iff (y) \subset (x)$. For instance in $\mathbf{Z}$, $2 \mid 6$ and $(6) \subset (2)$. We don't have $(2) \subset (6)$ since $2 \in (2)$ but $2 \notin (6)$. The successive divisibility relations $2 \mid 4 \mid 8 \mid 16 \mid \cdots$ correspond to the descending containment relations $(2) \supset (4) \supset (8) \supset (16) \supset \cdots$.

When does $(a) = (b)$? That is equivalent to $a \mid b$ and $b \mid a$, so $b = ac$ and $a = bd$ for some $c, d \in R$, which implies $b = bdc$ and $a = acd$. If this common ideal is not (0) and R is an *integral domain*, then $1 = dc$ and $1 = cd$, so c and d are invertible. Thus $a = bu$ where $u = d$ is invertible. Conversely, if $a = bu$ where u is invertible then $(a) = aR = buR = bR = (b)$, so we have shown that in an integral domain, a generator of a principal ideal is determined up to multiplication by a unit.

Here is the most important property of ideals in $\mathbf{Z}$ and $F[T]$, where F is a field.

Theorem 2.1. *In* $\mathbf{Z}$ *and* $F[T]$ *for every field* F*, all ideals are principal.*

Proof. Let I be an ideal in $\mathbf{Z}$ or $F[T]$. If $I = \{0\}$, then $I = (0)$ is principal. Let $I \neq (0)$. We have division with remainder in $\mathbf{Z}$ and $F[T]$ and will give similar proofs in both rings, side by side. *Learn this proof.*

Let $a \in I - \{0\}$ with $\lvert a\rvert$ minimal. So $(a) \subset I$. To show $I \subset (a)$, pick $b \in I$. Write $b = aq + r$ with $0 \leq r < \lvert a\rvert$. So $r = b - aq \in I$. By the minimality of $\lvert a\rvert$, $r = 0$. So $b = aq \in (a)$.	Let $f \in I - \{0\}$ with $\deg f$ minimal. So $(f) \subset I$. To show $I \subset (f)$, pick $g \in I$. Write $g = fq + r$ with $r = 0$ or $\deg r < \deg f$. So $r = g - fq \in I$. By the minimality of $\deg f$, $r = 0$. So $g = fq \in (f)$. $\square$

Example 2.2. In $\mathbf{Z}$, consider the finitely generated ideal

$$(6, 9, 15) = 6\mathbf{Z} + 9\mathbf{Z} + 15\mathbf{Z}.$$

This ideal must be principal, and in fact it is $3\mathbf{Z}$. To check the containment one way, since $6, 9, 15 \in 3\mathbf{Z}$ we get $6\mathbf{Z} + 9\mathbf{Z} + 15\mathbf{Z} \subset 3\mathbf{Z}$, and since $3 = -6 + 9 \in 6\mathbf{Z} + 9\mathbf{Z} + 15\mathbf{Z}$ we have $3\mathbf{Z} \subset 6\mathbf{Z} + 9\mathbf{Z} + 15\mathbf{Z}$. So the ideal $(6, 9, 15)$ is the principal ideal (3).

Remark 2.3. To check two finitely generated ideals $(r_1, \ldots, r_m)$ and $(r'_1, \ldots, r'_n)$ are equal, it is necessary and sufficient to check

$$r_1, \ldots, r_m \in (r'_1, \ldots, r'_n) \qquad \text{and} \qquad r'_1, \ldots, r'_n \in (r_1, \ldots, r_m).$$

For instance, to see in $\mathbf{Z}$ that $(6, 9, 15) = (3)$ we can observe that $6, 9, 15 \in (3)$ and $3 = -6 + 9 \in (6, 9, 15)$.

Example 2.4. For $\alpha \in \mathbf{C}$, let

$$I_\alpha = \{f(T) \in \mathbf{Q}[T] : f(\alpha) = 0\}.$$

This is an ideal in $\mathbf{Q}[T]$ (check!), so $I_\alpha = (h)$ for some $h \in \mathbf{Q}[T]$. Maybe the only polynomial in $\mathbf{Q}[T]$ that vanishes at α is 0 (*e.g.*, $\alpha = \pi = 3.1415\ldots$, which is transcendental). If there's some nonzero polynomial in $\mathbf{Q}[T]$ with α as a root then $I_\alpha \neq (0)$, so $h \neq 0$. The condition $I_\alpha = (h)$ means "for all $f \in \mathbf{Q}[T]$, $f(\alpha) = 0$ if and only if $h \mid f$". Note the similarity to orders in group theory: for $g \in G$, $\{n \in \mathbf{Z} : g^n = e\}$ is a subgroup of $\mathbf{Z}$ so it is $m\mathbf{Z}$ for some $m \in \mathbf{Z}$ with $m \geq 0$ (m is the order of g, if g has finite order): $g^n = e$ if and only if $m \mid n$.

Example 2.5. Which $f(T) \in \mathbf{R}[T]$ satisfy $f(i) = 0$? The set $I = \{f \in \mathbf{R}[T] : f(i) = 0\}$ forms an ideal in $\mathbf{R}[T]$ (check!) One such polynomial is $T^2 + 1$, so $(T^2 + 1) \subset I$. Let's show $I = (T^2 + 1)$. We know I is principal, say $I = (h)$. Then

$$T^2 + 1 \in (h) \Rightarrow h \mid T^2 + 1,$$

so $h = c$ or $h = c(T^2+1)$ for some $c \in \mathbf{R}^\times$. That means $(h) = (c) = (1)$ or $(h) = (T^2+1)$, but the former is impossible since the constant polynomial 1 is not in I. So $I = (h) = (T^2 + 1)$.

In Theorem 2.1 it is important that F is a field: if A is an integral domain and every ideal in $A[T]$is principal then A is a field. This is proved later in Theorem 6.12.

3. Ideals = Kernels, Quotient Rings

If $f\colon R \to S$ is a ring homomorphism, then $\ker f = \{r \in R : f(r) = 0\}$ is an ideal in R:

(1) it is an additive subgroup of R since f is an additive homomorphism.
(2) if $f(x) = 0$ and $r \in R$, then $rx \in \ker f$ since

$$f(rx) = f(r)f(x) = f(r) \cdot 0 = 0.$$

Not only is every kernel of a ring homomorphism defined on R an ideal in R, but all ideals in R arise in this way for some ring homomorphism out of R. Let's see some examples before proving this.

Example 3.1. For $m \in \mathbf{Z}$, the ideal $m\mathbf{Z}$ in $\mathbf{Z}$ is the kernel of the reduction homomorphism $\mathbf{Z} \to \mathbf{Z}/(m)$.

Example 3.2. For $\alpha \in \mathbf{C}$, the set $\{f \in \mathbf{Q}[T] : f(\alpha) = 0\}$ is the kernel of the evaluation-at-α homomorphism $\mathbf{Q}[T] \to \mathbf{C}$ where $f(T) \mapsto f(\alpha)$.

Example 3.3. For rings R and S, the ideals $R \times \{0\}$ and $\{0\} \times S$ in $R \times S$ are the kernels of the projection homomorphisms $R \times S \to S$ given by $(r, s) \mapsto s$ and $R \times S \to R$ given by $(r, s) \mapsto r$.

Theorem 3.4. *Every ideal in a ring R is the kernel of some ring homomorphism out of R.*

Proof. Since I is an additive subgroup we have the additive quotient group (of cosets)

$$R/I = \{r + I : r \in R\}.$$

Denote $r + I$ as $\overline{r}$. Under addition of cosets, the identity is $\overline{0}$ and the inverse of $\overline{r}$ is $\overline{-r}$. Define *multiplication* on R/I by

$$\overline{r} \cdot \overline{r'} = \overline{rr'}$$

for $\overline{r}, \overline{r'} \in R/I$. We need to check that this is well-defined: say $\overline{r}_1 = \overline{r}_2$ and $\overline{r'_1} = \overline{r'_2}$. Then $r_1 - r_2 = x \in I$ and $r'_1 - r'_2 = y \in I$. So to show $\overline{r_1 r'_1} = \overline{r_2 r'_2}$,

$$\begin{aligned} r_1 r'_1 - r_2 r'_2 &= (r_1 - r_2 + r_2)r'_1 - r_2 r'_2 \\ &= (r_1 - r_2)r_1 + r_2(r'_1 - r'_2) \\ &= xr'_1 + r_2 y \\ &\in I + I = I. \end{aligned}$$

Checking the rest of the conditions to have R/I be a ring is left to you.

The reduction mapping $R \to R/I$ by $r \mapsto \overline{r} = r + I$ is not just an additive group homomorphism but a ring homomorphism too. Indeed,

$$\overline{r_1 + r_2} = \overline{r}_1 + \overline{r}_2,\ \overline{r_1 r_2} = \overline{r}_1 \overline{r}_2,\ \overline{1} = \text{multiplicative identity in } R/I$$

The kernel of $R \to R/I$ is

$$\left\{r \in R : \overline{r} = \overline{0}\right\} = \{r : r + I = I\} = I,$$

so we have constructed an example of a ring homomorphism out of R with prescribed kernel I. This is analogous to the role of the canonical reduction homomorphism $G \to G/N$ in group theory that proves every normal subgroup N of a group G is the kernel of some group homomorphism out of G. □

Definition 3.5. For an ideal I in R, we call the ring R/I constructed in the above proof the *quotient ring* of R modulo I.

To be clear about what the ring R/I is, it is the additive quotient group R/I (treating R and I as additive groups) that is made into a ring by multiplying coset representatives, which is well-defined because I is an ideal.

Example 3.6. When $R = \mathbf{Z}$ and $I = m\mathbf{Z} = (m)$, $R/I = \mathbf{Z}/(m)$ is the usual ring of integers mod m.

Example 3.7. For a ring R, $R/(0) = R$ and $R/(1) = R/R = \{\overline{0}\}$ is the zero ring. So working modulo 0 changes nothing (congruence mod 0 is ordinary quality), while working modulo 1 collapses everything together.[1]

We'll see more interesting examples of quotient rings in the next section.

Remark 3.8. The additive quotient group $\mathbf{R}/\mathbf{Z}$, which is isomorphic to the circle group S^1, is *not* a ring in any reasonable way: $\mathbf{Z}$ is a subgroup of $\mathbf{R}$, not an ideal of $\mathbf{R}$ (the only ideals in $\mathbf{R}$ are (0) and $\mathbf{R}$), and multiplication on $\mathbf{R}/\mathbf{Z}$ doesn't make sense using representatives. Example, $1/2 = 5/2$ and $1/3 = 4/3$ in $\mathbf{R}/\mathbf{Z}$, but $(1/2)\cdot(1/3) \neq (5/2)\cdot(4/3)$ in $\mathbf{R}/\mathbf{Z}$ since $20/6 - 1/6 \notin \mathbf{Z}$.

4. The quotient is isomorphic to the image

In group theory, if $\varphi\colon G \to H$ is a group homomorphism with kernel N then φ is injective if and only if N is trivial, and $G/N \cong \varphi(G)$ as groups by $gN \mapsto \varphi(g)$. These results carry over to ring homomorphisms, using similar proofs.

Theorem 4.1. *If $\varphi\colon R \to S$ is a homomorphism of commutative rings with kernel I, then φ is injective if and only if $I = \{0\}$, and $R/I \cong \varphi(R)$ as rings by $\overline{r} \mapsto \varphi(r)$.*

Proof. Since φ is additive we have $\varphi(0) = 0$ (look at $\varphi(0) + \varphi(0) = \varphi(0+0) = \varphi(0)$ and subtract $\varphi(0)$ from both sides), so if φ is injective the only solution of $\varphi(r) = 0$ is $r = 0$. So when φ is injective, $I = \{0\}$.

Conversely, if $I = \{0\}$ then whenever $\varphi(x) = \varphi(y)$ we can say $\varphi(x-y) = \varphi(x) - \varphi(y) = 0$, so $x - y \in I = \{0\}$, so $x = y$. Thus φ is injective. (This proof, which only uses additivity properties of φ, is essentially the same as the proof in group theory that a group homomorphism is injective if and only if its kernel is trivial.)

Now assume $\varphi\colon R \to S$ is a ring homomorphism. We define a function $\overline{\varphi}\colon R/I \to S$ by

$$\overline{\varphi}(r + I) = \varphi(r).$$

This is well-defined: if $r + I = r' + I$ then $r = r' + x$ for some $x \in I$, so $\varphi(r) = \varphi(r' + x) = \varphi(r') + \varphi(x) = \varphi(r')$. Then the fact that φ is a ring homomorphism will imply $\overline{\varphi}$ is a ring homomorphism. For all r_1 and r_2 in R we have

$$\overline{\varphi}((r_1 + I) + (r_2 + I)) = \overline{\varphi}(r_1 + r_2 + I) = \varphi(r_1 + r_2)$$

and

$$\overline{\varphi}(r_1 + I) + \overline{\varphi}(r_2 + I) = \varphi(r_1) + \varphi(r_2)$$

so from $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$ we get that $\overline{\varphi}$ is additive. Multiplicativity of $\overline{\varphi}$ is shown in the same way: for all r_1 and r_2 in R,

$$\overline{\varphi}((r_1 + I)(r_2 + I)) = \overline{\varphi}(r_1 r_2 + I) = \varphi(r_1 r_2)$$

and

$$\overline{\varphi}(r_1 + I)\overline{\varphi}(r_2 + I) = \varphi(r_1)\varphi(r_2),$$

so from $\varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2)$ the mapping $\overline{\varphi}$ is multiplicative. Finally, $\overline{\varphi}(1 + I) = \varphi(1) = 1$.

Next we show $\overline{\varphi}\colon R/I \to S$ is injective. That is equivalent to showing its kernel is zero: if $\overline{\varphi}(r + I) = 0$ then $\varphi(r) = 0$ so $r \in I$, and thus $r + I$ is zero in R/I.

[1] In analysis, the additive group $\mathbf{R}/\mathbf{Z}$ is sometimes called "the real numbers mod 1", but that terminology is *not* related to what we're doing here: $\mathbf{R}/(1) = \mathbf{R}/\mathbf{R} = \{\overline{0}\}$ is a one-element ring while $\mathbf{R}/\mathbf{Z}$ is an infinite group with representatives in $[0, 1)$.

Finally, since $\overline{\varphi}(R/I) = \varphi(R)$, the injective homomorphism $\overline{\varphi}\colon R/I \to S$ has image $\varphi(R)$, so shrinking the target ring from S to $\varphi(R)$ we get a ring isomorphism (a bijective ring homomorphism) $R/I \to \varphi(R)$ using the function $\overline{\varphi}$. □

Example 4.2. Evaluation at 0 is a ring homomorphism $\mathbf{R}[T] \to \mathbf{R}$ that has kernel (T) and image $\mathbf{R}$ (look at the effect of evaluation on constant polynomials to see it is surjective!), so $\mathbf{R}[T]/(T) \cong \mathbf{R}$. By similar reasoning, for every ring A we have $A[T]/(T) \cong A$.

Example 4.3. Fix a real number c. Evaluation at c is a ring homomorphism $\mathbf{R}[T] \to \mathbf{R}$ that has kernel $(T-c)$ and image $\mathbf{R}$ (as in the previous example, the effect of this homomorphism on constant polynomials shows each real number is a value), so $\mathbf{R}[T]/(T-c) \cong \mathbf{R}$. In the same way, for every ring A we have $A[T]/(T-a) \cong A$ for all $a \in A$.

The way the two isomorphisms in the previous examples work on the congruence class of a particular polynomial is not the same (unless the polynomial is constant). Under evaluation at 0 we have $2T + 3 \bmod T$ corresponding to 3, while under evaluation at 1 we have $2T + 3 \bmod T - 1$ corresponding to 5.

Example 4.4. Evaluation at 0 is a ring homomorphism $\mathbf{Q}[T] \mapsto \mathbf{R}$ with kernel $T\mathbf{Q}[T] = (T)$ and image $\mathbf{Q}$, so $\mathbf{Q}[T]/(T) \cong \mathbf{Q}$. (Watch out: the image of this homomorphism is *not* $\mathbf{R}$, so we don't get an isomorphism from $\mathbf{Q}[T]/(T)$ to $\mathbf{R}$, but rather from $\mathbf{Q}[T]/(T)$ to $\mathbf{Q}$.)

Example 4.5. What is the ring $\mathbf{Q}[T]/(T^2)$? Modulo T^2, each polynomial in $\mathbf{Q}[T]$ is congruent to a unique polynomial of the form $a + bT$ for $a, b \in \mathbf{Q}$. In $\mathbf{Q}[T]/(T^2)$, T is not 0 but T^2 is 0. (This is analogous to $\mathbf{Z}/(9)$, where $3 \neq 0$ and $3^2 = 0$.) Therefore $\mathbf{Q}[T]/(T^2)$ consists of elements $a + b\overline{T}$ where $\overline{T} \neq 0$ and $\overline{T}^2 = 0$. Addition and multiplication in $\mathbf{Q}[T]/(T^2)$ is described by the formulas

$$(a + b\overline{T}) + (c + d\overline{T}) = (a + c) + (b + d)\overline{T}, \quad (a + b\overline{T})(c + d\overline{T}) = ac + (ad + bc)\overline{T}.$$

Example 4.6. Evaluation at i is a ring homomorphism $\mathbf{R}[T] \to \mathbf{C}$ that is surjective ($a+bi$ is the image of $a+bT$) and its kernel is (T^2+1), so we get a ring isomorphism $\mathbf{R}[T]/(T^2+1) \to \mathbf{C}$ by $f(T) \bmod T^2 + 1 \mapsto f(i)$. Coset representatives in $\mathbf{R}[T]/(T^2 + 1)$ can be chosen uniquely as polynomials of the form $a + bT$, and the addition and multiplication of these representatives in $\mathbf{R}[T]/(T^2+1)$ behaves exactly like addition and multiplication of complex numbers $a + bi$. The idea of using $\mathbf{R}[T]/(T^2 + 1)$ as a rigorous definition of the complex numbers goes back to Cauchy in 1847 [2], [3].

Example 4.7. Evaluation at $\sqrt[3]{2}$ is a ring homomorphism $\mathbf{Q}[T] \to \mathbf{R}$ whose kernel is $(T^3 - 2)$ and whose image is $\mathbf{Q}[\sqrt[3]{2}]$, so $\mathbf{Q}[T]/(T^3 - 2) \cong \mathbf{Q}[\sqrt[3]{2}]$.

5. Ideals of Polynomials

In geometry, ideals often – but not always – arise as the functions *vanishing* on a subset of some space. Let's look at some ideals of polynomials defined in this way.

Example 5.1. In $\mathbf{R}[X]$, the ideal

$$I = (X) = \{Xg(X) : g(X) \in \mathbf{R}[X]\}$$

is the set of polynomials in $\mathbf{R}[X]$ vanishing at 0.

Example 5.2. In $\mathbf{R}[X]$, the ideal

$$(X^2 + 1) = \left\{(X^2 + 1)g(X) : g(X) \in \mathbf{R}[X]\right\}$$

is $\{f(X) \in \mathbf{R}[X] : f(i) = 0\}$, which is the polynomials in $\mathbf{R}[X]$ that vanish at i.

Example 5.3. Let

$$I = \{f(X,Y) \in \mathbf{R}[X,Y] : f(0,0) = 0\} = \left\{ \sum_{i,j} a_{ij} X^i Y^j : a_{00} = 0 \right\}.$$

Elements of I look like

$$aX + bY + cX^2 + dXY + eY^2 + \cdots + fX^5Y^2 + \cdots.$$

These are the polynomials in $\mathbf{R}[X,Y]$ vanishing at $(0,0)$. We can write

$$I = \{Xg(X,Y) + Yh(X,Y) : g(X,Y), h(X,Y) \in \mathbf{R}[X,Y]\} = (X,Y).$$

We claim that I is *not* a principal ideal. The proof is by contradiction. Suppose $I = (k)$ for some polynomial $k = k(X,Y)$. Since X and Y are examples of elements of I, if we had such k then $k\ell = X$ and $km = Y$ for some polynomials ℓ and m in $\mathbf{R}[X,Y]$. This can only happen if k is a nonzero constant, but I contains no nonzero constants. Thus I is *not* principal.

Example 5.4. For a point $(a,b) \in \mathbf{R}^2$, let

$$I_{a,b} = \{f \in \mathbf{R}[X,Y] : f(a,b) = 0\}.$$

This ideal equals $(X-a, Y-b)$. To see why, since $X-a$ and $Y-b$ are in $I_{a,b}$ we have $(X-a, Y-b) \subset I_{a,b}$. To prove $I_{a,b} \subset (X-a, Y-b)$, here are two methods:

- Use the (finite!) Taylor expansion of polynomials at (a,b): each $f \in \mathbf{R}[X,Y]$ can be written as

  $$f(X,Y) = f(a,b) + \text{polynomial in } X-a, Y-b \text{ with no constant term},$$

  so when $f(a,b) = 0$, we have

  $$f(X,Y) \in (X-a, Y-b).$$

- In the ring $\mathbf{R}[X,Y]/(X-a, Y-b)$ we have $X \equiv a$ and $Y \equiv b$, so a polynomial expression in X and Y with real coefficients is congruent mod $(X-a, Y-b)$ to the same polynomial expression in a and b, or in other words $f(X,Y) \equiv f(a,b)$ mod $(X-a, Y-b)$. Thus when $f \in I_{a,b}$, meaning $f(a,b) = 0$, we get $f(X,Y) \in (X-a, Y-b)$.

Here is an incorrect proof that $I_{a,b} \subset (X-a, Y-b)$: if $f(X,Y) \in I_{a,b}$ then $f(X,b) \in \mathbf{R}[X]$ with a root at $X = a$ and $f(a,Y) \in \mathbf{R}[Y]$ with a root at $Y = b$, so $f(X,b) \in (X-a)$ in $\mathbf{R}[X]$ and $f(a,Y) \in (Y-b)$ in $\mathbf{R}[Y]$. That does not prove $f(X,Y) \in (X-a, Y-b)$ since it only uses ideals in the single-variable polynomial rings $\mathbf{R}[X]$ and $\mathbf{R}[Y]$ without making a link with the ideal $(X-a, Y-b)$ in $\mathbf{R}[X,Y]$: $(X-a)$ in $\mathbf{R}[X]$ is smaller than $(X-a)$ in $\mathbf{R}[X,Y]$, since the first $(X-a)$ only involves polynomials in X. Remember, as Remark 1.3 points out, that $(X-a, Y-b)$ is *not* the union of $(X-a)$ and $(Y-b)$ in $\mathbf{R}[X,Y]$.

The ideal $I_{a,b}$ is the kernel of the evaluation homomorphism $\mathbf{R}[X,Y] \to \mathbf{R}$ at (a,b), where $f(X,Y) \mapsto f(a,b)$. This ideal is non-principal by the same reasoning as in Example 5.3, which is that special case $a = 0$, $b = 0$.

Example 5.5. Consider the polynomials in $\mathbf{R}[X,Y]$ vanishing on the y-axis:

$$I = \{f \in \mathbf{R}[X,Y] : f(0,y) = 0 \text{ for all } y \in \mathbf{R}\}.$$

See Figure 1. Since $X \in I$,

$$(X) = \{X \cdot g(X,Y) : g \in \mathbf{R}[X,Y]\} \subset I.$$

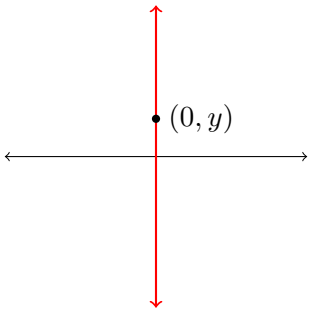

FIGURE 1. Solutions to $x = 0$.

In fact $(X) = I$. To show this, write each $f \in I$ in the form

$$f(X, Y) = h(Y) + X \cdot g(X, Y),$$

where $h(Y) \in \mathbf{R}[Y]$ is the "X-free" part of f. Then $f(0, Y) = h(Y)$, so $h(y) = 0$ for all $y \in \mathbf{R}$. The only polynomial in $\mathbf{R}[Y]$ with infinitely many roots is 0, so $h(Y) = 0$, so $f = Xg(X) \in (X)$.

Remark 5.6. Context matters with notation: the ideal (X) in $\mathbf{R}[X, Y]$ is not the same as the ideal (X) in $\mathbf{R}[X]$.

Example 5.7. Consider the polynomials in $\mathbf{R}[X, Y]$ vanishing on the parabola $y = x^2$:

$$\begin{aligned} I &= \left\{f \in \mathbf{R}[X, Y] : f(x, y) = 0 \text{ when } x, y \in \mathbf{R},\ y = x^2\right\} \\ &= \left\{f \in \mathbf{R}[X, Y] : f(x, x^2) = 0 \text{ for all } x \in \mathbf{R}\right\}. \end{aligned}$$

See Figure 2.

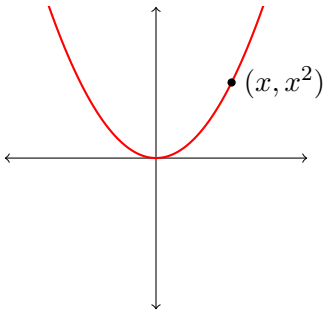

FIGURE 2. Solutions to $y = x^2$.

One polynomial in I is $Y - X^2$, so $(Y - X^2) \subset I$. In fact $I = (Y - X^2)$. To show this, pick $f(X, Y) \in I$. In the ring $\mathbf{R}[X, Y]/(Y - X^2)$ we have $Y \equiv X^2$ so $f(X, Y) \equiv f(X, X^2)$, so $f(X, Y) - f(X, X^2) \in (Y - X^2)$. The polynomial $f(X, X^2) \in \mathbf{R}[X]$ vanishes at each $x \in \mathbf{R}$, so $f(X, X^2) = 0$ in $\mathbf{R}[X]$. Therefore $f(X, Y) \in (Y - X^2)$.

Starting with the inclusion of points on a curve in the plane

$$\{(0, 0)\}, \{(2, 4)\} \subset \left\{(x, y) : y = x^2\right\} \subset \mathbf{R}^2,$$

passing to the ideal of polynomials vanishing on these sets reverses all inclusions:

$$(X,Y),(X-2,Y-4) \supset (Y-X^2) \supset (0).$$

It's easy to see algebraically that $(Y-X^2) \subset (X,Y)$ since $Y-X^2 \in (X,Y)$. While it's obvious *geometrically* that $(2,4)$ lies on the curve $y=x^2$, to check algebraically that $(Y-X^2) \subset (X-2,Y-4)$ can look tedious by comparison:

$$\begin{aligned} Y-X^2 &= Y-4+4-(X-2+2)^2 \\ &= Y-4+4-(X-2)^2-2\cdot 2(X-2)-4 \\ &= (Y-4)-(X-2)^2-4(X-2) \\ &\in (X-2,Y-4). \end{aligned}$$

6. Prime and Maximal Ideals

The rings whose behavior is closest to what is taught in high school algebra are integral domains and fields. It's important to know when a quotient ring R/I is an integral domain or a field, and such ideals I have special names.

Definition 6.1. An ideal $I \subset R$ is called a *prime* ideal if the quotient ring R/I is an integral domain. We call I a *maximal* ideal if the quotient ring R/I is a field.

Typically prime ideals are written as P and Q, while maximal ideals are written as M. Since the creators of ideal theory were German, we often follow their lead and write prime and maximal ideals using gothic fonts: $\mathfrak{p}$ and $\mathfrak{q}$ for prime ideals and $\mathfrak{m}$ for maximal ideals.

Example 6.2. In $\mathbf{Z}$, all ideals are $m\mathbf{Z}$ for $m \geq 0$ by Theorem 2.1. Furthermore, $\mathbf{Z}/(m)$ is an integral domain exactly when $m=0$ and $m=p$ is a prime number, and $\mathbf{Z}/(m)$ is a field exactly when $m=p$ is a prime number: if $a \not\equiv 0 \bmod p$ for a prime p then $\gcd(a,p)=1$ since p is prime, so $ax+py=1$ for some $x,y \in \mathbf{Z}$ and thus $ax \equiv 1 \bmod p$, so all nonzero elements of $\mathbf{Z}/(p)$ are units. When $m>1$ is not prime, a factor of m between 1 and m reduces to a zero divisor in $\mathbf{Z}/(m)$, so $\mathbf{Z}/(m)$ is not a field. So the prime ideals in $\mathbf{Z}$ are (0) and (p) for prime numbers p and the maximal ideals in $\mathbf{Z}$ are (p) for prime numbers p: the nonzero prime ideals in $\mathbf{Z}$ are maximal.

Similar reasoning shows that in $F[X]$ for a field F, the prime ideals are (0) and $(\pi(X))$ for irreducible $\pi(X) \in F[X]$ and the maximal ideals are $(\pi(X))$ for irreducible $\pi(X) \in F[X]$.

Example 6.3. In $\mathbf{Q}$ the only ideals are (0) and (1), with (0) being a maximal and prime ideal.

Example 6.4. The ideal (X) in $\mathbf{R}[X]$ is a maximal ideal since $\mathbf{R}[X]/(X) \cong \mathbf{R}$ (use evaluation at 0) and $\mathbf{R}$ is a field, while the ideal (X) in $\mathbf{R}[X,Y]$ is a prime ideal that is not a maximal ideal since $\mathbf{R}[X,Y]/(X) \cong \mathbf{R}[Y]$ (substitute 0 for X and view $\mathbf{R}[X,Y]$ as $\mathbf{R}[X][Y]$ and use Example 4.2) and $\mathbf{R}[Y]$ is an integral domain but not a field.

Example 6.5. The ideal $(Y-X^2)$ in $\mathbf{R}[X,Y]$ is prime and not maximal: the substitution homomorphism $\mathbf{R}[X,Y] \to \mathbf{R}[X]$ sending every $f(X,Y)$ to $f(X,X^2)$, or equivalently the evaluation homomorphism $\mathbf{R}[X][Y] \to \mathbf{R}[X]$ where $Y \mapsto X^2$, is surjective with kernel $(Y-X^2)$ by Example 5.7, so $\mathbf{R}[X,Y]/(Y-X^2) \cong \mathbf{R}[X]$, which is an integral domain but not a field.

Here are a few simple ways the terminology of prime and maximal ideals works.

- Since $R/(0) \cong R$, the ideal (0) in R is prime if and only if R is an integral domain and the ideal (0) in R is maximal if and only if R is a field.
- Every field is an integral domain, so *every maximal ideal is a prime ideal*: if R/I is a field then R/I is an integral domain. The converse is false, *e.g.*, (0) is a prime ideal in $\mathbf{Z}$ but not a maximal ideal and (Y) is a prime ideal in $\mathbf{R}[X, Y]$ but not a maximal ideal.
- The zero ring is *not* considered to be an integral domain or a field, since in an integral domain or field $1 \neq 0$ by definition. For a ring R, the quotient ring $R/(1)$ is the zero ring, so the ideal (1) is not considered to be a prime ideal or a maximal ideal: prime and maximal ideals are always *proper* ideals (not the whole ring).

Theorem 6.6. *An ideal I in R is prime if and only if $I \neq R$ and for all $a, b \in R$ the condition $ab \in I$ implies $a \in I$ or $b \in I$. An ideal I is maximal if and only if $I \neq R$ and for ideals J such that $I \subset J \subset R$, we have $J = I$ or $J = R$.*

This theorem explains the terminology "maximal": a maximal ideal is one that is truly *maximal* among all proper ideals of the ring.

Proof. To say R/I is an integral domain is the same as saying $R/I \neq \{\overline{0}\}$ and in R/I, if $\overline{a}\overline{b} = \overline{0}$, then $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$. This is equivalent to saying $I \neq R$ and if $ab \in I$ then $a \in I$ or $b \in I$, so those properties are equivalent to I being a prime ideal.

Now suppose R/I is a field and J is an ideal with $I \subset J \subset R$. To prove $J = I$ or $J = R$, assume $J \neq I$. We will show J contains 1, so $J = R$. Let $j \in J - I$, so in R/I we have $j \not\equiv 0 \bmod I$. Since R/I is a field, there is a $k \in R$ such that $jk \equiv 1 \bmod I$, so $jk = 1 + x$ for some $x \in I$. Thus $1 = jk - x$. Since $j \in J$ we have $jk \in J$, and since $x \in I \subset J$ we have $1 = jk - x \in J$. Thus $J = R$.

Conversely, suppose that I is a maximal ideal: it is proper ideal of R such that the only ideals J satisfying $I \subset J \subset R$ are $J = I$ or $J = R$. To prove R/I is a field, pick $a \neq 0$ in R/I. We will show a has an inverse in R/I. Consider the sum $I + Ra = \{x + ra : x \in I, r \in R\}$. This is an ideal in R (check!), it contains I (use $r = 0$), and it contains a (use $x = 0$ and $r = 1$), so the ideal $I + Ra$ is larger than I. Therefore $I + Ra = R$. That implies $1 = x + ra$ for some $x \in I$ and $r \in R$, so $ra \equiv 1 \bmod I$, and thus $a \bmod I$ has an inverse. □

Remark 6.7. In math, the word "or" allows for both options to happen: the condition "$xy = 0$ implies $x = 0$ or $y = 0$" in an integral domain absolutely includes the possibility that x and y are both 0. Just think about the integers: $xy = 0$ in $\mathbf{Z}$ tells us x is 0 or y is 0 or (just for emphasis!) both are 0, with the "both" bit being part of what "or" means so we don't make that explicit. Integral domains are supposed to mimic that situation, so an ideal $\mathfrak{p}$ is prime when it is a proper ideal and $ab \in \mathfrak{p}$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$ or (just for emphasis!) both a and b are in $\mathfrak{p}$.

We mentioned above that while all maximal ideals are prime (because all fields are integral domains), not all prime ideals in a ring have to be maximal. The following theorem gives a set of conditions that are sufficient for all nonzero prime ideals to be maximal, so the maximal ideals are the nonzero prime ideals.

Theorem 6.8. *If R is an integral domain in which all ideals are principal then every nonzero prime ideal in R is maximal.*

It is crucial that we refer in the theorem to *nonzero* prime ideals: by Example 6.2, in $\mathbf{Z}$ the nonzero prime ideals iare maximal but the zero ideal is prime and not maximal. The zero ideal is maximal only in a field.

Proof. Write a nonzero prime ideal of R as (p) for some $p \in R$ (the ideal is principal by hypothesis). To prove (p) is maximal, let I be an ideal with $(p) \subset I \subset R$. We will show $I = (p)$ or $I = R$.

By hypothesis, $I = (a)$ for some $a \in R$. Then the condition $(p) \subset I$ says $(p) \subset (a)$, so $p \in (a)$. Thus $p = ab$ for some $b \in R$, so $ab \equiv 0 \bmod (p)$. Since (p) is a prime ideal, $R/(p)$ is an integral domain and therefore $a \equiv 0 \bmod (p)$ or $b \equiv 0 \bmod (p)$. We will show one of these cases leads to $(a) = (p)$ and the other leads to $(a) = R$.

If $a \equiv 0 \bmod (p)$ then $a = pa'$ for some $a' \in R$, so $p = ab = pa'b$. Since R is an integral domain, $1 = a'b$, so b is a unit. Thus $(a) = (ab) = (p)$.

If $b \equiv 0 \bmod (p)$ then $b = pb'$ for some $b' \in R$, so $p = ab = pab'$. As before we can cancel p, getting $1 = ab'$, so a is a unit. Thus $(a) = (1) = R$. □

Definition 6.9. An integral domain in which all ideals are principal is called a *principal ideal domain*, which is abbreviated to PID.

Example 6.10. The rings $\mathbf{Z}$ and $F[T]$ for a field F are PIDs by Theorem 2.1.

Theorem 6.8 says that in a PID, every nonzero prime ideal is maximal. The converse has many counterexamples: an integral domain in which all nonzero prime ideals are maximal can have nonprincipal ideals.

Example 6.11. In $\mathbf{Z}[\sqrt{5}] = \{a + b\sqrt{5} : a, b \in \mathbf{Z}\}$, it can be shown that all nonzero prime ideals of $\mathbf{Z}[\sqrt{5}]$ are maximal, but the ideal $(2, 1+\sqrt{5})$ in $\mathbf{Z}[\sqrt{5}]$ is nonprincipal. More generally, if d is an integer that is not a perfect square then all nonzero prime ideals in $\mathbf{Z}[\sqrt{d}]$ turn out to be maximal, and when $d \equiv 1 \bmod 4$ (such as $d = 5, 13, 17, 21, -3, -7, -11$ and -15) the ideal $(2, 1 + \sqrt{d})$ in $\mathbf{Z}[\sqrt{d}]$ turns out to be nonprincipal.

When F is a field, $F[T]$ is a PID by Theorem 2.1. The next theorem is a converse result.

Theorem 6.12. *If A is a commutative ring such that $A[T]$ is a* PID *then A is a field.*

Proof. To begin, A is an integral domain since A is a subring of $A[T]$ and a subring of an integral domain is an integral domain.

To prove A is a field when $A[T]$ is a PID, we will give two proofs. The second proof is more conceptual than the first.

For the first proof, pick $a \neq 0$ in A. We want to show a is a unit in A. By hypothesis, the ideal (a, T) in $A[T]$ is principal, say $(a, T) = (f(T))$. Then $f(T) \neq 0$, $f(T) \mid a$, and $f(T) \mid T$. Write the first divisibility condition as $a = f(T)g(T)$ for some $g(T) \in A[T]$, so $g(T) \neq 0$. Since A is an integral domain we have $\deg(fg) = \deg f + \deg g$, so $\deg f + \deg g = 0$. Thus $f(T)$ is constant, say $f(T) = c$. Then $c \mid T$, so $T = ch(T)$ for an $h(T) \in A[T]$. Comparing leading coefficients (or coefficients of T) on both sides, we get $c \in A^\times$, so $(f(T)) = (c) = A[T]$. Thus $(a, T) = (f(T)) = A[T]$, so 1 is in (a, T): $1 = au(T) + Tv(T)$ for some $u(T)$ and $v(T)$ in $A[T]$. Comparing constant terms on both sides shows $a \in A^\times$, so we have shown all nonzero elements of A are units in A: A is a field.

For the second proof that A is a field, the ideal (T) in $A[T]$ is prime since $A[T]/(T) \cong A$ is an integral domain, so by Theorem 6.8 with $R = A[T]$ the ideal (T) in $A[T]$ is maximal. Therefore $A[T]/(T)$ is a field, so A is a field. □

The second proof of Theorem 6.12 is more slick than the first, but it doesn't give us an *example* of a nonprincipal ideal in $A[T]$ when A is an integral domain that is not a field. The first proof tells us some examples: the ideal (a, T) in $A[T]$ is not principal if $a \in A$ is

not 0 or a unit. For example, $(2, T)$ is not principal in $\mathbf{Z}[T]$ and (X, Y) is not principal in $\mathbf{R}[X, Y] = (\mathbf{R}[X])[Y]$. It is worthwhile writing out the proof that $(2, T)$ is nonprincipal in $\mathbf{Z}[T]$ using the method shown in the proof of Theorem 6.12.

Theorems 2.1 and 6.12 tell us that A is a field if and only if $A[T]$ is a PID. Without assuming A is an integral domain, having all ideals of $A[T]$ be principal is *not* equivalent to A being a field. For example, all ideals in $(\mathbf{Z}/(6))[T]$ turn out to be principal but $\mathbf{Z}/(6)$ is not a field. When A is an arbitrary commutative ring, all ideals in $A[T]$ are principal if and only if A is a finite product of fields.[2] A proof of this is in Pete Clark's answer at https://math.stackexchange.com/questions/361258/.

The last thing we will show about maximal ideals is that every nonzero ring contains a maximal ideal, and thus also a prime ideal (since all maximal ideals are prime). Some rings have only one maximal ideal (like (0) in $\mathbf{Q}$), and in some rings it may be hard to describe the maximal ideals, but at least one exists.[3] The proof of this uses Zorn's lemma, a fundamental set-theoretic result that is equivalent to the axiom of choice. This is usually the first time students meet Zorn's lemma in algebra. Here is the statement of Zorn's lemma.

> *Zorn's lemma*: If S is a nonempty partially ordered set and every totally ordered subset has an upper bound in S then S has a maximal element m: there is an $m \in S$ such that $x \leq m$ for all $x \in S$ to which m is comparable.

Theorem 6.13. *Every nonzero commutative ring R contains a maximal ideal.*

Proof. We will use Zorn's lemma. Consider the set of all proper ideals in R:

$$S = \{I \subset R : I \text{ ideal}, \ I \neq R\}.$$

The set S is nonempty since $(0) \in S$. Partially order S by inclusion; i.e. $I \leq J$ means that $I \subseteq J$. Suppose we have a *totally ordered* subset $\{I_\alpha\}_{\alpha \in A}$. Let

$$I = \bigcup_{\alpha \in A} I_\alpha.$$

This is an ideal: say $x, y \in I$. Then $x \in I_\alpha$ and $y \in I_\beta$ for some $\alpha, \beta \in A$. Either $I_\alpha \subseteq I_\beta$ or $I_\beta \subseteq I_\alpha$ because our subset of S is totally ordered. Then $x + y \in I_\beta \subseteq I$ or $x + y \in I_\alpha \subseteq I$. Either way we get $x + y \in I$. If $x \in I$, so $x \in I_\alpha$ for some α, and $r \in R$, then $rx \in I_\alpha \subseteq I$. This shows I is an ideal in R.

The ideal I is proper: if $I = R$, then $1 \in I$, so $1 \in I_\alpha$ for some α, which is impossible as each I_α is proper. So $I \in S$ and $I_\alpha \subseteq I$ for all $\alpha \in A$. We've shown every totally ordered subset of S has an upper bound in S. So by Zorn's lemma, S contains a maximal element. A maximal element of S is, by definition, a proper ideal in R that is not contained in a proper ideal other than itself, and such an ideal is maximal ideal by Theorem 6.6. □

Note that the upper bounds constructed on totally ordered subsets of S are typically *not* the maximal elements coming from Zorn's lemma. That is, the justification to apply Zorn's lemma is a completely separate task from actually applying Zorn's lemma and seeing what can be said about a maximal element. For example, if S is the set of all proper ideals of $\mathbf{Z}$, partially ordered by inclusion, then the totally ordered subset of ideals $\{12^k\mathbf{Z} : k \geq 1\}$ has upper bounds in $\mathbf{Z}$ such as $12\mathbf{Z}$ or $4\mathbf{Z}$, which are not maximal elements of S.

[2]This fits the example $A = \mathbf{Z}/(6)$ because $\mathbf{Z}/(6) \cong \mathbf{Z}/(2) \times \mathbf{Z}/(3)$ by the Chinese remainder theorem.

[3]In contrast, a nontrivial group need not have a maximal proper subgroup. For instance, every proper subgroup of $\mathbf{Q}$ is contained in a larger proper subgroup, so there is no maximal proper subgroup of $\mathbf{Q}$.

Corollary 6.14. *In a nonzero commutative ring, every proper ideal is contained in a maximal ideal.*

Proof. Let R be a nonzero commutative ring and J be a proper ideal in R. We want to show there is a maximal ideal M in R such that $J \subset M \subset R$.

The quotient ring R/J is nonzero, so by Theorem 6.13 it contains a maximal ideal, say $\overline{M}$. The composite of the reduction maps $R \to R/J \to (R/J)/\overline{M}$ is a surjective ring homomorphism. Let M denote the kernel, so by Theorem 4.1 there is an induced ring isomorphism $R/M \cong (R/J)/\overline{M}$. Therefore R/M is a field, so M is maximal in R. Since elements of J vanish in $(R/J)/\overline{M}$, $J \subset M$. □

Corollary 6.14 could be proved with Zorn's lemma by modifying the proof of Theorem 6.13: show there is a maximal element in the set S of proper ideals in R that contain J. The purpose of proving Corollary 6.14 in the way we did above is to illustrate how passage to a quotient ring can let us reduce a question about general ideals to the special case of the ideal (0). This is a very useful method in algebra.

We will use Corollary 6.14 near the end of the next section to create the nonstandard real numbers.

7. The real numbers as a quotient ring

As an application of quotient rings, in this section we will construct $\mathbf{R}$ from $\mathbf{Q}$. Before we do this in Definition 7.6, the only numbers we will use are rational.

Every real number should be a limit of a sequence of rational numbers, which suggests we could define a real number as a sequence of rational numbers that (intuitively) has that real number as a limit. At the same time, different sequences in $\mathbf{Q}$ can have the same limit (consider $(0, 0, 0, 0, \ldots)$, $(1, 1/2, 1/3, 1/4, \ldots)$, and $(1/4, -1/9, 1/16, -1/25, \ldots)$, so we need to decide when two sequences in $\mathbf{Q}$ should converge to the same real number without mentioning real numbers. There are two tasks: (i) describe the sequences in $\mathbf{Q}$ that ought to converge in $\mathbf{R}$ without using limits (since a limit may not be rational) and (ii) describe when two such sequences in $\mathbf{Q}$ have the same limit, so they should be the same "real number."

Definition 7.1. A sequence $\mathbf{x} = \{x_k\}$ in $\mathbf{Q}$ is called *Cauchy* if for all rational $\varepsilon > 0$ there is a $K \in \mathbf{Z}^+$ such that $k, \ell \geq K \Longrightarrow |x_k - x_\ell| \leq \varepsilon$.

The intuition behind this definition is that in a Cauchy sequence the terms don't just get consecutively close ($x_k - x_{k-1}$ tends to 0), but uniformly close: $x_k - x_\ell$ is small for all large k and ℓ. The partial sums of the harmonic series, $H_k = 1+1/2+\cdots+1/k$, get consecutively close but diverge, so consecutive closeness can not be used in the definition of a Cauchy sequence. Every convergent sequence is a Cauchy sequence,[4] and Cauchy sequences are the sequences that "want" to converge even if there is no actual limit yet.

Lemma 7.2. *If $\mathbf{x} = \{x_k\}$ is a Cauchy sequence in $\mathbf{Q}$ then it is bounded: there is a rational number $b > 0$ such that $|x_k| \leq b$ for all k.*

Proof. In the definition of $\mathbf{x}$ being a Cauchy sequence let $\varepsilon = 1$. Then there is some $K \in \mathbf{Z}^+$ such that $k, \ell \geq K \Longrightarrow |x_k - x_\ell| \leq 1$. In particular, if $k \geq K$ then $|x_k - x_K| \leq 1$, so

$$k \geq K \Longrightarrow |x_k| = |x_k - x_K + x_K| \leq |x_k - x_K| + |x_K| \leq 1 + |x_K|.$$

Therefore we can use for b the maximum of $|x_1|, |x_2|, \ldots, |x_{K-1}|$ and $1 + |x_K|$. □

[4] If $x_k \to x$ then for all rational $\varepsilon > 0$ there is a K such that $k \geq K \Rightarrow |x - x_k| \leq \varepsilon/2$, so $k, \ell \geq K \Rightarrow |x_k - x_\ell| = |(x_k - x) + (x - x_\ell)| \leq |x_k - x| + |x - x_\ell| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$.

Denote by C the set of all Cauchy sequences in $\mathbf{Q}$, and by S the set of all sequences in $\mathbf{Q}$ (Cauchy or not), so $C \subset S$ and S is a commutative ring with componentwise operations, additive identity $\mathbf{0} = (0, 0, 0, \ldots)$, and multiplicative identity $\mathbf{1} = (1, 1, 1, \ldots)$. Constant sequences are Cauchy, so $\mathbf{Q}$ embeds into C by identifying $r \in \mathbf{Q}$ with the constant sequence $(r, r, r, \ldots)$. The next theorem implies C is a subring of S.

Theorem 7.3. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are Cauchy sequences in* $\mathbf{Q}$ *then* $\mathbf{x} \pm \mathbf{y}$ *and* $\mathbf{xy}$ *are also Cauchy.*

Proof. Pick a rational $\varepsilon > 0$.

To prove the sequence $\mathbf{x} + \mathbf{y} = \{x_k + y_k\}$ is Cauchy consider the inequality

$$|(x_k + y_k) - (x_\ell + y_\ell)| = |x_k - x_\ell + y_k - y_\ell| \leq |x_k - x_\ell| + |y_k - y_\ell|.$$

This suggests applying the definition of a Cauchy sequence with $\varepsilon/2$ instead of ε: there is some $K \in \mathbf{Z}^+$ such that $k, \ell \geq K \Longrightarrow |x_k - x_\ell| \leq \varepsilon/2$ and $|y_k - y_\ell| \leq \varepsilon/2$.[5] Then

$$k, \ell \geq K \Longrightarrow |(x_k + y_k) - (x_\ell + y_\ell)| \leq |x_k - x_\ell| + |y_k - y_\ell| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The proof that $\mathbf{x} - \mathbf{y}$ is Cauchy is nearly the same, and details are left to the reader.

Proving $\mathbf{xy}$ is Cauchy is more subtle. Consider the inequality

$$|x_k y_k - x_\ell y_\ell| = |(x_k - x_\ell)y_k + (y_k - y_\ell)x_\ell| \leq |x_k - x_\ell||y_k| + |y_k - y_\ell||x_\ell|. \tag{7.1}$$

The sequences $\mathbf{x}$ and $\mathbf{y}$ are bounded by Lemma 7.2, so using a common bound for both there is some rational $b > 0$ such that $|x_k| \leq b$ and $|y_k| \leq b$ for all k. Then by (7.1)

$$|x_k y_k - x_\ell y_\ell| \leq |x_k - x_\ell| b + |y_k - y_\ell| b.$$

That suggests using $\varepsilon/(2b)$ in place of ε in the definition of Cauchy sequences: there is some K such that $k, \ell \geq K \Longrightarrow |x_k - x_\ell| \leq \varepsilon/(2b)$ and $|y_k - y_\ell| \leq \varepsilon/(2b)$. Then

$$k, \ell \geq K \Longrightarrow |x_k y_k - x_\ell y_\ell| \leq \frac{\varepsilon}{2b} b + \frac{\varepsilon}{2b} b = \varepsilon. \qquad \square$$

It is intuitively clear that two convergent sequences have the same limit if and only if their difference sequence tends to 0. That motivates the next definition.

Definition 7.4. A sequence of rational numbers $\mathbf{x} = \{x_k\}$ is called a *null sequence* if $x_k \to 0$: for all rational $\varepsilon > 0$ there is a K such that for $k \geq K$ we have $|x_k| \leq \varepsilon$.

Let N denote the set of all null sequences in $\mathbf{Q}$.

Theorem 7.5. *The set* N *is a proper ideal in* C*.*

Proof. First we check $N \subset C$. For $\mathbf{x}$ in N and a rational $\varepsilon > 0$, use $\varepsilon/2$ in the definition of a null sequence: there is some K such that for all $k \geq K$ we have $|x_k| \leq \varepsilon/2$. Then for all $k, \ell \geq K$ we have $|x_k - x_\ell| \leq |x_k| + |x_\ell| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$, so $\{x_k\}$ is Cauchy.

The proof that the sum and difference of two null sequences is a null sequence uses a similar $\varepsilon/2$ argument, and is left to the reader.

Suppose $\mathbf{x} \in N$ and $\mathbf{y} \in C$. To prove $\mathbf{xy} \in N$, by Lemma 7.2 the sequence $\mathbf{y}$ is bounded, say $|y_k| \leq b$ for some rational $b > 0$ and all k. Then $|x_k y_k| \leq |x_k| b$, so if for a rational $\varepsilon > 0$ we use ε/b in place of ε in the definition of $\mathbf{x}$ being a null sequence it follows from the upper bound on $|x_k y_k|$ that $\mathbf{xy}$ is a null sequence.

The ideal N is not all of C since it does not contain, for instance, the constant sequences $(r, r, r, \ldots)$ where $r \in \mathbf{Q}^\times$. $\square$

[5]Strictly speaking the choice of K at first depends on the choice of sequence $\mathbf{x}$ or $\mathbf{y}$, but by using the larger of the two K's we can use one K for both.

Since C is a commutative ring and N is a proper ideal in C, C/N is a nonzero commutative ring using addition and multiplication of coset representatives.

Definition 7.6. The *real numbers* $\mathbf{R}$ are defined to be C/N: Cauchy sequences in $\mathbf{Q}$ modulo sequences in $\mathbf{Q}$ that tend to 0.

By the construction of quotient rings $\mathbf{R}$ is a commutative ring. The composition $\mathbf{Q} \to C \to C/N$, where the first mapping is $r \mapsto (r, r, r, \ldots)$ and the second is reduction, is a ring homomorphism. It is injective since $(r, r, r, \ldots) \in N$ only if $r = 0$. Thus we can view $\mathbf{Q}$ as a subfield of $\mathbf{R}$.

Theorem 7.7. *The ring* $\mathbf{R}$ *is a field.*

Proof. We want to prove each nonzero element of $\mathbf{R}$ has an inverse: if $\mathbf{x}$ is a Cauchy sequence in $\mathbf{Q}$ that is not a null sequence we will find a Cauchy sequence $\mathbf{y}$ such that $\mathbf{xy} \equiv \mathbf{1} \bmod N$, or equivalently $x_k y_k - 1 \to 0$. In fact we'll show for all large k that $x_k \neq 0$ and we can use $y_k = 1/x_k$ for large k.

Claim: a Cauchy sequence in $\mathbf{Q}$ that does not tend to 0 is eventually bounded away from 0: there is some rational $c > 0$ and $k_0 \in \mathbf{Z}^+$ such that $|x_k| \geq c$ for all $k \geq k_0$.

The proof of the claim will need the Cauchy property, as a general sequence not tending to 0 does not have to be eventually bounded away from 0: consider $1, 0, 1, 0, 1, 0, \ldots$.

To prove the claim we prove its contrapositive: a Cauchy sequence $\mathbf{x}$ that is not eventually bounded away from 0 must be a null sequence. Not being eventually bounded away from 0 means it is *not true* that there is a rational $c > 0$ and a k_0 such that $k \geq k_0 \Longrightarrow |x_k| \geq c$. So for all rational $\varepsilon > 0$ there is no k_0 such that $k \geq k_0 \Longrightarrow |x_k| \geq \varepsilon$,[6] hence for all rational $\varepsilon > 0$ and all k_0 there is some $k \geq k_0$ such that $|x_k| < \varepsilon$. Starting with one k_0 and $k \geq k_0$ such that $|x_k| < \varepsilon$, repeatedly picking a new k_0 that exceeds the previously chosen k and then a new k greater than or equal to the new k_0 so that $|x_k| < \varepsilon$, we get for each rational $\varepsilon > 0$ that $|x_k| < \varepsilon$ for infinitely many k. Taking $\varepsilon = 1, 1/2, 1/3, \ldots$, this implies that a *subsequence* of $\mathbf{x}$ tends to 0. The Cauchy property will let us bootstrap this to show the whole sequence $\mathbf{x}$ tends to 0, *i.e.*, $\mathbf{x}$ is a null sequence.

To prove $x_k \to 0$ means for all rational $\varepsilon > 0$ we want to show there is some K such that $k \geq K \Longrightarrow |x_k| \leq \varepsilon$. Since $\mathbf{x}$ is Cauchy, there is a K such that $k, \ell \geq K \Longrightarrow |x_k - x_\ell| \leq \varepsilon/2$. From the previous paragraph with $\varepsilon/2$ in place of ε, there are infinitely many indices $k_1 < k_2 < k_3 < \cdots$ such that $|x_{k_i}| \leq \varepsilon/2$. Eventually the k_i are greater than or equal to K, and using such k_i in the role of ℓ from the Cauchy condition we get

$$k \geq K \Longrightarrow |x_k| = |x_k - x_{k_i} + x_{k_i}| \leq |x_k - x_{k_i}| + |x_{k_i}| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

That completes the proof of (the contrapositive of) the claim.

Using c and k_0 as in the claim, for $k \geq k_0$ we have $|x_k| \geq c > 0$, so $x_k \neq 0$. Define a sequence of rational numbers $\mathbf{y}$ by

$$y_k = \begin{cases} 1/x_k, & \text{if } k \geq k_0, \\ 1, & \text{if } k < k_0. \end{cases}$$

Then for $k, \ell \geq k_0$ we have

$$|y_k - y_\ell| = \left| \frac{1}{x_k} - \frac{1}{x_\ell} \right| = \frac{|x_k - x_\ell|}{|x_k||x_\ell|} \leq \frac{|x_k - x_\ell|}{c^2},$$

[6] We change the letter c to ε for psychological purposes.

and from $\mathbf{x}$ being Cauchy this bound implies $\mathbf{y}$ is Cauchy: for all rational $\varepsilon > 0$ there is a K such that $k, \ell \geq K \Rightarrow |x_k - x_\ell| \leq \varepsilon c^2$, so $k, \ell \geq \max(K, k_0) \Rightarrow |y_k - y_\ell| \leq (\varepsilon c^2)/c^2 = \varepsilon$.

Since $x_k y_k = 1$ for $k \geq k_0$, the difference $\mathbf{xy} - \mathbf{1}$ has k-th component 0 for all $k \geq k_0$. A sequence whose terms eventually all equal 0 is in N, so $\mathbf{xy} - \mathbf{1} \in N$ and therefore in $\mathbf{R} = C/N$ we get $\mathbf{xy} \equiv \mathbf{1} \bmod N$. □

This theorem proves that C/N is a field, so N is a maximal ideal in C.

There is more that should be done: define an ordering on $\mathbf{R}$ (that is, define positive and negative) in terms of representative rational Cauchy sequences, show every real number is a limit of rational numbers, and show every Cauchy sequence *of real numbers* converges (this is the completeness property: Cauchy = convergent for sequences in $\mathbf{R}$). Details of these properties are at the end of [5, §3, Chap. IX], from which our treatment is adapted.

How does the construction of $\mathbf{R}$ from $\mathbf{Q}$ as a quotient ring compare to what is done in analysis books? There are two common ways of defining $\mathbf{R}$ from $\mathbf{Q}$: Dedekind cuts and equivalence classes of Cauchy sequences of rational numbers. Dedekind cuts are formalizations of subsets of $\mathbf{Q}$ like $\{r \in \mathbf{Q} : r < x\}$ for real x that make no direct reference to x itself. The idea is that each real number is characterized by the rationals that are less than it. Dedekind cuts are used in [1, §8.6], [7, §2, Chap. 1], [8, §6, Chap. 1], and [9, App., Chap. 1], and get rather ugly for multiplication because defining that operation requires many cases and proving properties with that definition is tedious. The other method, using Cauchy sequences in $\mathbf{Q}$, is in [10, Chap. 2] and [11, Chap. 5]. It uses an equivalence relation on C:

$$\{x_k\} \sim \{y_k\} \Longleftrightarrow x_k - y_k \to 0.$$

It is not hard to check this is an equivalence relation: $\{x_k\} \sim \{x_k\}$, if $\{x_k\} \sim \{y_k\}$ then $\{y_k\} \sim \{x_k\}$, and if $\{x_k\} \sim \{y_k\}$ and $\{y_k\} \sim \{z_k\}$ then $\{x_k\} \sim \{z_k\}$. The real numbers are defined as equivalence classes of Cauchy sequences in $\mathbf{Q}$ for the relation $\sim$. This is the same as our C/N since Cauchy sequences in $\mathbf{Q}$ are equivalent for $\sim$ precisely when their difference is in N, so $\{x_k\} \sim \{y_k\} \Leftrightarrow \{x_k\} \equiv \{y_k\} \bmod N$. Equivalence classes for $\sim$ are the same as cosets in C/N. The sum and product of equivalence classes are $\overline{\{x_k\}} + \overline{\{y_k\}} = \overline{\{x_k + y_k\}}$ and $\overline{\{x_k\}} \cdot \overline{\{y_k\}} = \overline{\{x_k y_k\}}$. Checking these are well-defined amounts to an argument like the one used to prove addition and multiplication in a quotient ring using coset representatives are well-defined; the case of multiplication requires an additional step essentially equivalent to proving N is an ideal.

What happens if we consider the construction analogous to C/N using real numbers instead of rational numbers: Cauchy sequences in $\mathbf{R}$ modulo null sequences in $\mathbf{R}$? Because all real Cauchy sequences have a real limit, this construction essentially gives us $\mathbf{R}$ back. But there is something interesting that can be done with the product ring of *all* real sequences

$$\mathbf{R}^\infty = \prod_{k \geq 1} \mathbf{R} = \{(a_1, a_2, a_3, \ldots) : a_k \in \mathbf{R}\},$$

which at first looks too big to be useful (there are so many non-Cauchy sequences!).

For each $n \geq 1$ the ideal $V_n = \{\mathbf{a} \in \mathbf{R}^\infty : a_n = 0\}$ in $\mathbf{R}^\infty$ is principal, generated by $(1, 1, \ldots, 1, 0, 1, \ldots)$, which is 0 in the nth component and 1 elsewhere, with $\mathbf{R}^\infty/V_n \cong \mathbf{R}$ by projection $\mathbf{R}^\infty \to \mathbf{R}$ onto the nth component. Thus each ideal V_n is maximal and the quotient ring $\mathbf{R}^\infty/V_n$ is not anything new since it is isomorphic to $\mathbf{R}$.

Consider a new ideal in $\mathbf{R}^\infty$: the sequences in $\mathbf{R}$ whose terms are 0 beyond some point:

$$V = \{\mathbf{a} \in \mathbf{R}^\infty : a_k = 0 \text{ for all large enough } k\}.$$

This is an ideal, and it is proper since it doesn't contain $(1, 1, 1, \ldots)$. By Corollary 6.14, the proper ideal V of $\mathbf{R}^\infty$ is contained in some maximal ideal M.[7] We have $V \not\subset V_n$ for each n since the sequence that is 1 in the nth component and 0 elsewhere is in V but not V_n. Since $V \subset M$, M is not one of the ideals V_n. The next theorem shows each maximal ideal of $\mathbf{R}^\infty$ other than $V_1, V_2, V_3, \ldots$ can be described in the way we defined M.

Theorem 7.8. *A maximal ideal of $\mathbf{R}^\infty$ that is not some V_n must contain V.*

Proof. Let M be a maximal ideal of $\mathbf{R}^\infty$ with $M \neq V_n$ for all $n \geq 1$. Let $\mathbf{x}_n = (\ldots, 0, 1, 0, \ldots)$ be 1 in the nth component and 0 everywhere else and $\mathbf{y}_n = (\ldots, 1, 0, 1, \ldots)$ be 0 in the nth component and 1 everywhere else. Then V_n is the principal ideal $(\mathbf{y}_n) = \mathbf{y}_n\mathbf{R}^\infty$.

Since $\mathbf{x}_n\mathbf{y}_n = \mathbf{0} \in M$ and $\mathbf{R}^\infty/M$ is a field, for each $n \geq 1$ we have $\mathbf{x}_n \in M$ or $\mathbf{y}_n \in M$. If $\mathbf{y}_n \in M$ then $V_n = (\mathbf{y}_n) \subset M$, so $M = V_n$ since V_n is maximal. That is a contradiction, so $\mathbf{x}_n \in M$ for all n. That implies each sequence that is nonzero in only one component is in M, so by adding finitely many such sequences together we get $V \subset M$ (why?). □

While there are many choices for M, none of them can be described in a concrete way. The field $\mathbf{R}^\infty/M$, which is usually denoted ${}^*\mathbf{R}$, has a unique ordering and is called a model for the *nonstandard real numbers*. It contains $\mathbf{R}$ (as the image of $\mathbf{R} \to \mathbf{R}^\infty \to \mathbf{R}^\infty/M$ where the first mapping is on the diagonal and the second is the canonical reduction map) and also contains infinitely large and infinitely small numbers. How does $\mathbf{R}^\infty/M$, up to field isomorphism, depend on the choice of M that contains V? This question is closely related to set theory: to say all choices of maximal ideal M containing V lead to isomorphic fields $\mathbf{R}^\infty/M$ is equivalent to the continuum hypothesis[8] [6]. The fields ${}^*\mathbf{R}$ and $\mathbf{R}$ are elementarily equivalent in the sense of model theory, and this is codified in the transfer principle. For more on ${}^*\mathbf{R}$ see [4, Chap. 12] and watch the YouTube video "Hyperreal Numbers" by blargoner.

References

[1] S. Abbott, "Understanding Analysis," 2nd ed., Springer-Verlag, 2015.

[2] A. L. Cauchy, Mémorie sur une nouvelle théorie des imaginaires, et sur les racines symboliques des équationes et des équivalences, *Comptes Rendus hebdomadaires des séances de l'Académie des Sciences de Paris* **24** (1847), 1120–1130. Also *Ouvres Complètes*, Sér. 1, Tome 10, pp. 312–323. URL https://www.biodiversitylibrary.org/item/21169#page/1130/mode/1up.

[3] A. L. Cauchy, Mémoire sur la théorie des équivalences algébriques, substituée à la théorie des imaginaires, pp. 87–110 in "Exercises d'analyse et de physique mathématique", Tome 4 (1847). Also *Ouvres Complètes*, Sér. 2, Tome 14, pp. 93–120. URL https://archive.org/details/117770570_004/page/n91/mode/2up.

[4] H.-D. Ebbinghaus *et al.*, "Numbers," Springer-Verlag, 1991.

[5] S. Lang, "Undergraduate Algebra," 2nd ed., Springer-Verlag, 1990.

[6] LCL, Ultrapowers and hyperreals, http://math.stackexchange.com/questions/719131 (version: 2014-03-20).

[7] C. C. Pugh, "Real Mathematical Analysis," 2nd ed., Springer-Verlag, 2015.

[8] K. Ross, "Elementary Analysis: The Theory of Calculus," 2nd ed., Springer-Verlag, 2013.

[9] W. Rudin, "Principles of Mathematical Analysis," 3rd ed., McGraw-Hill, 1976.

[10] R. S. Strichartz, "The Way of Analysis," Revised edition, Jones and Bartlett, 2000.

[11] T. Tao, "Analysis I, Volume 1," 3rd ed., Hindustan Book Agency, 2006.

[7] The ideal V is not a maximal ideal, or even a prime ideal, since $\mathbf{R}^\infty/V$ is not an integral domain: the sequences $(1, 0, 1, 0, 1, 0, \ldots)$ with alternating terms 1 and 0 and $(0, 1, 0, 1, 0, 1, \ldots)$ with alternating terms 0 and 1 are not in V but their product is, so in $\mathbf{R}^\infty/V$ their cosets are nonzero with product equal to 0.

[8] In the ring $\mathbf{C}[X]$ the quotients by all maximal ideals are isomorphic since $\mathbf{C}[X]/(X - a) \cong \mathbf{C}$ for all $a \in \mathbf{C}$.