Notes on Abstract Algebra

Adam Coffman

Contents

These Notes are compiled from classroom handouts for Math 453 and 553 at IPFW. They are not self-contained, but supplement the required texts, [D], [BB2]/[BB3]/[BB4], respectively. Some related books are [H] and [CLO].

1 Functions

Definition 1.1. Given sets S and T, suppose there is a subset $G \subseteq S \times T$ with the following properties:

- If $(s_1, t_1) \in G$ and $(s_2, t_2) \in G$ and $s_1 = s_2$ then $t_1 = t_2$;
- For each $s \in S$, there is an element $(s, t) \in G$.

Then for each $s \in S$, there is exactly one element $\alpha(s) \in T$ so that $(s, \alpha(s)) \in G$. This defines a function α, with domain S and target T, which can be denoted $\alpha : S \to T$.

Theorem 1.2. *Given $S \neq \emptyset$, and a function $\alpha : S \to T$, the following are equivalent:*

1. *For all s_1, $s_2 \in S$, if $s_1 \neq s_2$, then $\alpha(s_1) \neq \alpha(s_2)$ (α has the <u>one-to-one</u> property);*
2. *For any set C and any functions $\gamma : C \to S$, $\delta : C \to S$, if $\alpha \circ \gamma : C \to T$ and $\alpha \circ \delta : C \to T$ are the same function, then $\gamma = \delta$ (α has the <u>left cancellable</u> property);*
3. *There is a function $\beta : T \to S$ so that $\beta \circ \alpha : S \to S$ is equal to the identity function $\iota : S \to S$ (α has a <u>left inverse</u>).*

Proof. (3) $\implies$ (2): Given α, we assume there is a left inverse β as in (3). If γ and δ satisfy $\alpha \circ \gamma = \alpha \circ \delta$, then $\beta \circ (\alpha \circ \gamma) = \beta \circ (\alpha \circ \delta)$, so by the associative property of composition, $(\beta \circ \alpha) \circ \gamma = (\beta \circ \alpha) \circ \delta$. It follows that $\iota \circ \gamma = \iota \circ \delta$, and we get the conclusion $\gamma = \delta$, so (2) holds.

(2) $\implies$ (1): We give a proof of the contrapositive, assuming (1) is false, and then showing (2) is also false. Suppose there are some $s_1, s_2 \in S$ so that $s_1 \neq s_2$ but $\alpha(s_1) = \alpha(s_2)$. Let C be the two element set $\{0, 1\}$, and define the constant function $\gamma : C \to S$ by the graph $\{(0, s_1), (1, s_1))\}$ and a different constant function $\delta : C \to S$ by the graph $\{(0, s_2), (1, s_2)\}$. Then for any $x \in C$, $(\alpha \circ \gamma)(x) = \alpha(\gamma(x)) = \alpha(s_1)$, and $(\alpha \circ \delta)(x) = \alpha(\delta(x)) = \alpha(s_2)$. So, $\alpha \circ \gamma$ and $\alpha \circ \delta$ are both equal to the constant function $C \to T$ with value $\alpha(s_1) = \alpha(s_2)$. This means (2) is false, since $\alpha \circ \gamma = \alpha \circ \delta$ but $\gamma \neq \delta$.

(1) $\implies$ (3): Assuming (1) holds, consider the graph of α, the subset $G \subseteq S \times T$. The one-to-one property (1) can be re-phrased: if (s_1, t_1) and (s_2, t_2) are in G and $t_1 = t_2$ then $s_1 = s_2$.

We want to construct a function $\beta : T \to S$. First pick any $s_0 \in S$ (possible since S is non-empty). For any function α with graph G, each $t \in T$ falls into one of two possible types: given t, either there is at least one $s \in S$ so that $(s, t) \in G$ (t is type I), or there is no $s \in S$ so that $(s, t) \in G$ (t is type II). Define $H \subseteq T \times S$ as the union of the two sets:

$$H = \{(t, s) : t \text{ is type I, and } (s, t) \in G\} \cup \{(t, s_0) : t \text{ is type II}\}.$$

For each $t \in T$, there is some point $(t, s) \in H$. To show H is the graph of a function β, consider $(t_1, s_1) \in H$ and $(t_2, s_2) \in H$; we want to show that if $t_1 = t_2$, then $s_1 = s_2$. If t_1 is type I, then $(s_1, t_1) \in G$, and if $t_2 = t_1$, then t_2 is also type I (since, for example, $s = s_1$ satisfies $(s, t_2) = (s_1, t_1) \in G$), so $(s_2, t_2) \in G$. By the one-to-one property of G, $(s_1, t_1), (s_2, t_2) \in G$ and $t_1 = t_2$ implies $s_1 = s_2$. If t_1 is type II, then $(t_1, s_1) = (t_1, s_0)$, and if $t_2 = t_1$, then t_2 is also type II (since any $s \in S$ satisfying $(s, t_2) \in G$ would also satisfy $(s, t_1) \in G$, and there is no such s), so $(t_2, s_2) = (t_2, s_0)$, and $s_2 = s_0 = s_1$. We can conclude that the formula: $\beta(t) = s$, if t is type I with $\alpha(s) = t$, and $\beta(t) = s_0$, if t is type II, defines a function $\beta : T \to S$.

Considering the composite $\beta \circ \alpha$, pick any $s \in S$, so that $(s, \alpha(s)) \in G$. Then, let $t = \alpha(s)$; since $(s, t) \in G$, t is type I and $\beta(t)$ is, by definition, an element of S such that $(\beta(t), t) \in G$. By the one-to-one property of G, $(s, \alpha(s))$, $(\beta(t), t) \in G$ and $\alpha(s) = t$ implies $s = \beta(t)$. The conclusion is that $s = \beta(\alpha(s))$, so $\beta \circ \alpha = \iota$. ∎

Theorem 1.3. *Given a function $\alpha : S \to T$, the following are equivalent:*

1. *For all $t \in T$, there is some $s \in S$ so that $\alpha(s) = t$ (α has the onto property);*
2. *For any set C and any functions $\gamma : T \to C$, $\delta : T \to C$, if $\gamma \circ \alpha : S \to C$ and $\delta \circ \alpha : S \to C$ are the same function, then $\gamma = \delta$ (α has the right cancellable property);*
3. *There is a function $\beta : T \to S$ so that $\alpha \circ \beta : T \to T$ is equal to the identity function $\iota : T \to T$ (α has a right inverse).*

Proof. (3) $\implies$ (2): Given α, we assume there is a right inverse β as in (3). If γ and δ satisfy $\gamma \circ \alpha = \delta \circ \alpha$, then $(\gamma \circ \alpha) \circ \beta = (\delta \circ \alpha) \circ \beta$, so by the associative property of composition, $\gamma \circ (\alpha \circ \beta) = \delta \circ (\alpha \circ \beta)$. It follows that $\gamma \circ \iota = \delta \circ \iota$, and we get the conclusion $\gamma = \delta$, so (2) holds.

(2) $\implies$ (1): We give a proof of the contrapositive, assuming (1) is false, and then showing (2) is also false. Suppose there is some $t_0 \in T$ so that $\alpha(s) \neq t_0$ for all $s \in S$. Let C be the two element set $\{0, 1\}$, and define the constant function $\gamma : T \to C$ by $\gamma(t) = 0$ for all $t \in T$, and define a non-constant function $\delta : T \to C$, by $\delta(t) = 0$ for all $t \neq t_0$, and $\delta(t_0) = 1$. Then for any $s \in S$, $(\gamma \circ \alpha)(s) = \gamma(\alpha(s)) = 0$, and $(\delta \circ \alpha)(s) = \delta(\alpha(s)) = 0$, since $\alpha(s) \neq t_0$. So, $\gamma \circ \alpha$ and $\delta \circ \alpha$ are both equal to the constant function $S \to C$ with value 0. This means (2) is false, since $\gamma \circ \alpha = \delta \circ \alpha$ but $\gamma \neq \delta$.

(1) $\implies$ (3): Assuming (1) holds, consider the graph of α, the subset $G \subseteq S \times T$. We want to construct a function $\beta : T \to S$. For each $t \in T$, there is, by the onto property, at least one $s \in S$ such that $(s, t) \in G$, so define $\beta(t) \in S$ by choosing any such s. (**)

Considering the composite $\alpha \circ \beta$, for any $t \in T$, $(\alpha \circ \beta)(t) = \alpha(\beta(t)) = \alpha(s)$, where $\beta(t) = s$ was chosen so that $(s, t) \in G$. This means $t = \alpha(s)$, and the conclusion is that $t = \alpha(\beta(t))$, so $\alpha \circ \beta = \iota$.

■

(** footnote remark on this step) If T is an infinite set, this step in the proof requires that we make infinitely many choices, one choice for each $t \in T$. We will assume that we don't have any problem with doing that, but in some branches of mathematics, the ability to make an infinite sequence of choices isn't automatically assumed and needs to be taken as another hypothesis for the Theorem.

2 Binary operations

Definition 2.1. Given a set S, a binary operation on S is any function from $S \times S$ to S. The notation $(S, *)$ denotes a set S, together with $*$, a binary operation on S. For $x, y \in S$, and a binary operation $*$, the element $*((x, y))$ will be abbreviated $x * y$.

Definition 2.2. A binary operation $*$ on S is associative means: for all $x, y, z \in S$, $(x * y) * z = x * (y * z)$. It is commutative means: for all $x, y \in S$, $x * y = y * x$.

Definition 2.3. Given $(S, *)$, any element $e \in S$ such that $e * x = x * e = x$ for all $x \in S$ is called an identity element.

Proposition 2.4. *Given $(S, *)$, suppose there is an identity element $e \in S$. Then, the identity element is unique.* ■

Exercise 2.5. Given $(S, *)$, with an identity element e, if for all $x, y, z \in S$, $x*(y*z) = (x*z)*y$, then $*$ is commutative and associative. ■

The following problems show that the hypothesis $e \in S$ is necessary.

Exercise 2.6. Let S be the set of matrices,

$$S = \left\{ \begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} : x, y, z \in \mathbb{R} \right\},$$

and let $*$ be matrix multiplication. Show that the formula $A * (B * C) = (A * C) * B$ holds for all matrices $A, B, C \in S$. Show, by giving an explicit numerical example, that $*$ is not commutative on S. ■

Exercise 2.7. Using the same set S as the previous Exercise, replace the operation $*$ by $\odot$, where $A \odot B$ is defined by $A * B - B * A$. Is $\odot$ associative? Does the formula $A \odot (B \odot C) = (A \odot C) \odot B$ hold for all A, B, C? Is $\odot$ commutative? $\blacksquare$

Exercise 2.8. Give an example of a set and an operation $*$ where $x * (y * z) = (x * z) * y$ holds but $*$ is not associative. $\blacksquare$

Definition 2.9. Given $(S, *)$, and an identity element $e \in S$, and $x, y \in S$, y is a $*$-inverse for x means that $x * y = y * x = e$.

Note that $*$-inverse cannot be defined without an identity element, so in any statement asserting the existence of a $*$-inverse, it is assumed that there exists an identity element for the operation $*$.

Exercise 2.10. Given $(S, *)$, let e be an identity element. Then e has a $*$-inverse, and this inverse is unique. $\blacksquare$

Exercise 2.11. Given $(S, *)$, and $x \in S$, if $*$ is associative, and there exists a $*$-inverse for x, then that $*$-inverse for x is unique. $\blacksquare$

Notation 2.12. Usually it is more convenient to call a $*$-inverse just an "inverse," and if an element x has a unique inverse, it can be denoted x^{-1}. There may be some other abbreviations for certain operations; customarily a $+$-inverse of x is denoted $-x$.

Exercise 2.13. Given $(S, *)$, and $x, y \in S$, if $*$ is associative, and x and y both have $*$-inverses, then $x * y$ has a unique $*$-inverse, $y^{-1} * x^{-1}$. $\blacksquare$

If $*$ is not associative, $(S, *)$ could still have an identity element, but some elements could have more than one inverse.

Example 2.14. Let $S = \{e, a, b, c\}$, and define a binary operation $*$ by the table

$*$	e	a	b	c
e	e	a	b	c
a	a	a	e	e
b	b	e	a	e
c	c	a	e	b

Note that e is an identity element, and $*$ is not commutative ($a * c = e$, $c * a = a$). It is also not associative ($a * (a * b) = a * e = a$, $(a * a) * b = a * b = e$). The equations $x * b = e$, $b * x = e$ have two solutions ($x = a$ and $x = c$). So, b has an inverse, but b does not have a unique inverse. Since $c * b = e$, and there is only one solution of the equation $c * x = e$, c has a unique inverse, $c^{-1} = b$. Even though $a * c = e$, a is not an inverse for c because $c * a \neq e$.

Exercise 2.15. Given $(S, *)$, and $x \in S$, if $*$ is associative, and there exists a $*$-inverse for x, and $x * x = x$, then $x = e$. $\blacksquare$

The above condition that $*$ is associative cannot just be dropped — the conclusion may not follow if $*$ is not associative. In Example 2.14, a has a unique inverse ($a * b = b * a = e$), but $a * a = a$, and $a \neq e$.

Notation 2.16. Given $(S,*)$, $x \in S$, and a natural number $n \in \mathbb{N}$, define x^n to be x if $n = 1$, and then for $n > 1$, recursively define x^n to be $(x^{(n-1)}) * x$. If $(S,*)$ has an identity element $e \in S$, then x^0 is defined to be e. If the element $x \in S$ has a unique $*$-inverse $x^{-1} \in S$, and n is a negative integer, then x^n is defined to be to $(x^{-1})^{(-n)}$.

This notation is consistent with the appearance of $-1 \in \mathbb{Z}$ as an exponent in Notation 2.12. The "order of operations" convention should be familiar: $x * y^n$ means $x * (y^n)$, so that the exponentiation is done before the $*$ operation, and $x * y^n$ is, in general, not the same as $(x * y)^n$, where the parentheses indicate the $*$ is done first.

Again, the notation can change for certain operations; when the operation is $+$, the notation is customarily written as nx, for example, $4x = ((x + x) + x) + x$.

Theorem 2.17. *Given $(S,*)$, with an identity element e, if n is any integer, then $e^n = e$.*

Proof. For $n = 0$ and $n = 1$, the claim is obvious. Suppose, inductively, that $n > 1$, and that the claim is true for $n - 1$, which means $e^{n-1} = e$. Then $e^n = e^{n-1} * e$, by definition, and this equals $e * e$ by the inductive hypothesis. Then, by the definition of identity element, $e * e = e$, so $e^n = e$.

Since e is its own unique inverse, negative powers of e are well-defined, and if n is a negative integer, then $e^n = (e^{-1})^{(-n)} = e^{(-n)} = e$. ■

Exercise 2.18. Given $(S,*)$, suppose e is an identity element. If x and y are any elements of S, then $(x * y)^0 = x^0 * y^0$, and if n is any nonnegative integer, then $(x^0) * (x^n) = (x^n) * (x^0) = x^n$, and $e = x^0 = (x^0)^n = (x^n)^0$. If x has a unique inverse, then $(x^0) * (x^n) = (x^n) * (x^0) = x^n$ and $e = x^0 = (x^0)^n = (x^n)^0$ for all $n \in \mathbb{Z}$.

Hint. These are easy to prove — but it is worth remarking that these statements, and Theorem 2.17, do not require $*$ to be associative. ■

Lemma 2.19. *Given $(S,*)$, and $x, y \in S$, if $*$ is associative and $x*y = y*x$ then $x*y^n = y^n*x$ for all $n \in \mathbb{N}$. If, also, y has a $*$-inverse, then $x * y^n = y^n * x$ for all $n \in \mathbb{Z}$.*

Proof. Note that $*$ doesn't have to be a commutative operation: we're only assuming that x and y commute with each other and not necessarily with other elements of S.

The claim is clearly true for $n = 1$ (and does not even require $*$ to be associative).

For $n > 1$, assume that the statement has been proved for $n - 1$: the "inductive hypothesis" is that $x * y^{n-1} = y^{n-1} * x$, and we want to show that $x * y^n = y^n * x$.

Using the definition of nth power and the associative law, $x*y^n = x*(y^{n-1}*y) = (x*y^{n-1})*y$, and the inductive hypothesis says this equals $(y^{n-1} * x) * y$, which, using the associativity and the $x * y = y * x$ hypothesis, equals $y^{n-1} * (x * y) = y^{n-1} * (y * x) = (y^{n-1} * y) * x$, which by definition is $y^n * x$.

If $e \in S$ is an identity element, then $x*y^0 = y^0*x$, even if $*$ is not associative and $x*y \neq y*x$.

If y has an inverse, it is unique by Exercise 2.11, and the $n = -1$ claim is that if $x*y = y*x$, then $x * y^{-1} = y^{-1} * x$, which is proved by the following steps:

$$\begin{aligned}
x * y^{-1} &= (e * x) * y^{-1} = ((y^{-1} * y) * x) * y^{-1} = (y^{-1} * (y * x)) * y^{-1} \\
&= (y^{-1} * (x * y)) * y^{-1} = y^{-1} * ((x * y) * y^{-1}) = y^{-1} * (x * (y * y^{-1})) \\
&= y^{-1} * (x * e) = y^{-1} * x.
\end{aligned}$$

If n is any negative integer, then $x*y^n = x*(y^{-1})^{-n}$ by definition, and since $x*y^{-1} = y^{-1}*x$ (which was just proved), the first part of the Theorem applies to x, y^{-1}, and the positive integer $-n$, giving $x*(y^{-1})^{-n} = (y^{-1})^{-n}*x$, so $x*y^n = y^n*x$. ∎

The associativity is a necessary hypothesis. In Example 2.14, $b*c = c*b$, but $c*b^2 = c*a = a$, and $b^2*c = a*c = e$.

Theorem 2.20. *Given $(S,*)$, and $x, y \in S$, if $*$ is associative and $x*y = y*x$ then $(x*y)^n = x^n*y^n$ for all $n \in \mathbb{N}$. If, also, x and y have $*$-inverses, then $(x*y)^n = x^n*y^n$ holds for all $n \in \mathbb{Z}$.*

Proof. The statement is clearly true for $n = 1$ (and does not even require associativity or $x*y = y*x$).

For $n > 1$, assume that the statement has been proved for $n-1$: the "inductive hypothesis" is that $(x*y)^{n-1} = x^{n-1}*y^{n-1}$, and we want to show that $(x*y)^n = x^n*y^n$. By definition, $(x*y)^n = ((x*y)^{n-1})*(x*y)$, and this equals $(x^{n-1}*y^{n-1})*(x*y)$ by the inductive hypothesis. Then, by associativity, it equals $((x^{n-1}*y^{n-1})*x)*y = (x^{n-1}*(y^{n-1}*x))*y$. Here, Lemma 2.19 applies, which is where the $x*y = y*x$ hypothesis is needed, to give $(x^{n-1}*(x*y^{n-1}))*y$, which by associativity equals $((x^{n-1}*x)*y^{n-1})*y = (x^{n-1}*x)*(y^{n-1}*y)$, and by definition, equals $x^n y^n$.

The second claim allows n to be 0; since x and y have inverses, there must be an identity element e, and then Exercise 2.18 applies, even if x and y don't commute. To prove the second claim for negative n, x, y, and $x*y$ have unique inverses, by Exercises 2.11 and 2.13 (which also require associativity), and $(x*y)^n$ is defined to be $((x*y)^{-1})^{(-n)}$. By the $x*y = y*x$ hypothesis and Exercise 2.13, this is equal to $((y*x)^{-1})^{(-n)} = (x^{-1}*y^{-1})^{(-n)}$, which by the first claim of the Theorem, is equal to $(x^{-1})^{(-n)}*(y^{-1})^{(-n)} = x^n*y^n$. ∎

Corollary 2.21. *Given $(S,*)$, and $x, y \in S$, if $*$ is associative and y is a $*$-inverse of x, then y^n is a $*$-inverse of x^n and $x^{(-n)} = y^n$ for all $n \in \mathbb{Z}$.*

Proof. By the uniqueness of inverses for associative operations (Exercise 2.11), y is the unique inverse of x ($y = x^{-1}$) and x is the unique inverse of y ($x = y^{-1}$). Since x and y have inverses and commute by definition ($x*y = y*x = e$), Theorem 2.20 applies to give $x^n*y^n = (x*y)^n = e^n$, and $e^n = e$ by Theorem 2.17. Similarly, $y^n*x^n = (y*x)^n = e^n = e$. It can be concluded that $y^n = (x^n)^{-1} = (x^{-1})^n$. The proof of the second equality needs a few cases: the $n = 0$ case is easy, and if $n > 0$, then $x^{(-n)}$ is defined to be $(x^{-1})^{-(-n)} = y^n$. If $n < 0$, then y^n is defined to be $(y^{-1})^{(-n)} = x^{(-n)}$. ∎

The associativity is a necessary hypothesis for both the Theorem and the Corollary. In Example 2.14, $b*a = a*b = e$, but $b^2*a^2 = a*a = a \neq e = (b*a)^2$.

Theorem 2.22. *Given $(S,*)$, and $g \in S$, if $*$ is associative, then $(g^a)*(g^b) = g^{(a+b)}$ for any $a, b \in \mathbb{N}$. If g has a $*$-inverse, then $(g^a)*(g^b) = g^{(a+b)}$ for any $a, b \in \mathbb{Z}$.*

Proof. The claim is true for any a if $b = 1$, since $(g^a)*(g^1) = (g^a)*g = g^{(a+1)}$ by definition of nth power for $n = a+1$. Suppose, inductively, that $b > 1$, and that the claim is true for $b-1$, which means $(g^a)*(g^{(b-1)}) = g^{(a+(b-1))}$. Then $(g^a)*(g^b) = (g^a)*((g^{(b-1)})*g)$ by definition of bth power, which equals $((g^a)*(g^{(b-1)}))*g$ by associativity, and equals $(g^{(a+(b-1))})*g$ by the inductive hypothesis. Then the definition of nth power again shows this is equal to $g^{(a+(b-1))+1} = g^{(a+b)}$, which proves the claim for the positive integer b.

If $e \in S$ is an identity element, the claim for $a = 0$ or $b = 0$ was stated in Exercise 2.18.

If g has an inverse, it is unique by Exercise 2.11. The rest of the proof proceeds in several cases. Case 1. If a and b are both negative, then $g^a * g^b = ((g^{-1})^{-a}) * ((g^{-1})^{-b})$, and the first claim applies to the positive exponents $-a$, $-b$, and $g^{-1} \in S$, so $((g^{-1})^{-a}) * ((g^{-1})^{-b}) = (g^{-1})^{(-a)+(-b)} = g^{-((-a)+(-b))} = g^{a+b}$.

There are several remaining cases, when a and b have opposite signs. Case 2. $0 < -a = b$. $g^a * g^b = ((g^{-1})^{-a}) * g^b = ((g^{-1})^b) * g^b = e$, by Cor. 2.21. Then, $e = g^0 = g^{a+b}$ by definition of g^0. Case 3. $0 < a = -b$. $g^a * g^b = g^a * ((g^{-1})^{-b}) = g^a * ((g^{-1})^a) = e = g^{a+b}$. Case 4. $0 < -a < b$. $g^a * g^b = ((g^{-1})^{-a}) * ((g^{-a}) * (g^{a+b}))$, using the claim for the positive exponents $-a$, $b + a$, and $b = (-a) + (b + a)$. Then, by associativity, the product is equal to $(((g^{-1})^{-a}) * (g^{-a})) * (g^{a+b}) = e * (g^{a+b}) = g^{a+b}$. Case 5. $0 < b < -a$. $g^a * g^b = ((g^{-1})^{-a}) * g^b = (((g^{-1})^{-a-b}) * ((g^{-1})^b)) * g^b = ((g^{-1})^{-(a+b)}) * (((g^{-1})^b) * g^b) = (g^{a+b}) * e = g^{a+b}$. Case 6. $0 < a < -b$. $g^a * g^b = g^a * ((g^{-1})^{-b}) = g^a * (((g^{-1})^a) * ((g^{-1})^{(-a)+(-b)})) = (g^a * ((g^{-1})^a)) * ((g^{-1})^{-(a+b)}) = e * (g^{a+b}) = g^{a+b}$. Case 7. $0 < -b < a$. $g^a * g^b = ((g^{a+b}) * (g^{-b})) * ((g^{-1})^{-b}) = (g^{a+b}) * ((g^{-b}) * ((g^{-1})^{-b})) = (g^{a+b}) * e = g^{a+b}$. ∎

Example 2.23. The associativity hypothesis is necessary. Try, for example, $S = \mathbb{N}$, and let $*$ be the operation of integer exponentiation ($* = \hat{\ }$), so $m * n$ is m multiplied $(\cdot)$ by itself n times, and $*$ is not associative, for example, $(2*3)*2 = 64$, and $2*(3*2) = 512$. Theorem 2.22 doesn't work, where "3^3" is defined to be $(3*3)*3 = 27 \cdot 27 \cdot 27 = 19683$. This is the same as $(3^2)*(3^1)$, but not the same as $(3^1)*(3^2) = 3 * 27 = 3 \cdot 3 \cdot 3 \cdots 3 = 7625597484987$.

Theorem 2.24. *Given $(S, *)$, and $g \in S$, if $*$ is associative, then $g^{(mn)} = (g^m)^n$ for any $m, n \in \mathbb{N}$. If g has a $*$-inverse, then $g^{(mn)} = (g^m)^n$ for any $m, n \in \mathbb{Z}$.*

Proof. The claim is true for any m if $n = 1$, since $g^{(m \cdot 1)} = g^m = (g^m)^1$. Suppose, inductively, that $n > 1$, and that the claim is true for $n - 1$, which means $g^{(m(n-1))} = (g^m)^{(n-1)}$. Then $(g^m)^n = ((g^m)^{(n-1)}) * (g^m)$ by the definition of nth power, which equals $(g^{(m(n-1))}) * (g^m)$ by the inductive hypothesis. By Theorem 2.22 on adding exponents, $(g^{(m(n-1))}) * (g^m) = g^{(m(n-1))+m} = g^{mn}$, which proves the claim for the positive integer n.

If there is an identity element e, Exercise 2.18 applies to prove the second claim in the case that $m = 0$ or $n = 0$. It remains to check the cases where m or n is negative. If m is any integer, and n is positive, the above proof by induction will work even for negative m as long as g has an inverse, using Theorem 2.22 to add any integer exponents.

If m is positive, and n is negative, then mn is negative, so by definition, $g^{(mn)} = (g^{-1})^{-(mn)} = (g^{-1})^{(m(-n))}$, and since m and $-n$ are positive, the first claim gives $((g^{-1})^m)^{(-n)}$, and then Corollary 2.21 gives $((g^m)^{-1})^{(-n)} = (g^m)^n$. If m and n are both negative, then $g^{(mn)} = g^{(-m)(-n)} = (g^{(-m)})^{(-n)}$ by the first claim, and then Corollary 2.21 again gives $((g^m)^{-1})^{(-n)} = (g^m)^n$. ∎

The associativity is a necessary hypothesis here, too. In Example 2.14, $c^4 = ((c * c) * c) * c = (b * c) * c = e * c = c$, but $(c^2)^2 = (c * c) * (c * c) = b * b = a$.

3 Ideals in a Rng

Definition 3.1. A set S, together with two binary operations $+_S$ and $\cdot_S$, is a rng means that $+_S$ and $\cdot_S$ are associative, $+_S$ is commutative, there is an element 0_S which is an identity element for $(S, +_S)$, and every $x \in S$ has a $+_S$-inverse, denoted $-_S x$, or just $-x$. The two operations (abbreviated $+$ and $\cdot$) are related by the distributive laws: for all $a, b, c \in S$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

A rng S where $\cdot_S$ is commutative is called a commutative rng.

Given $x \in S$, a $+_S$-inverse $-x$ can be called the opposite of x in S; it is unique by Exercise 2.11. Subtraction in S is defined by $x - y = x -_S y = x +_S (-y) = x + (-y)$.

Definition 3.2. Let $(R, +_R, \cdot_R)$ be a rng. A set S is a subrng of R means:

- $S \subseteq R$, and
- There are operations $+_S$ and $\cdot_S$ such that $(S, +_S, \cdot_S)$ is a rng, and
- For all $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} +_R \mathbf{y} = \mathbf{x} +_S \mathbf{y}$, and
- For all $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} \cdot_R \mathbf{y} = \mathbf{x} \cdot_S \mathbf{y}$.

Theorem 3.3. *If S is a subrng of R, where R has zero element 0_R, then 0_R is an element of S, and is equal to the zero element of S.*

Proof. By the second part of Definition 3.2, S is a rng, so by Definition 3.1 applied to S, S contains a zero element $0_S \in S$. By the first part of Definition 3.2, $S \subseteq R$, which implies $0_S \in R$. By Definition 3.1 applied to S, $0_S +_S 0_S = 0_S$, and by Definition 3.2, $0_S +_R 0_S = 0_S +_S 0_S$. It follows that $0_S +_R 0_S = 0_S \in R$, and then Exercise 2.15, applied to the associative operation $+_R$ and the element $0_S \in R$, which has a $+_R$-inverse in R, implies $0_S = 0_R$. ■

Theorem 3.3 can be used in this way: if S is a set that does not contain 0_R as one of its elements, then S is not a subrng of R.

Theorem 3.4. *If S is a subrng of a rng R, then for every $\mathbf{w} \in S$, the opposite of $\mathbf{w}$ in S is the same as the opposite of $\mathbf{w}$ in R.*

Proof. Let $\mathbf{w}$ be an element of S; then $\mathbf{w} \in R$ because $S \subseteq R$.

First, we show that an additive inverse of $\mathbf{w}$ in S is also an additive inverse of $\mathbf{w}$ in R. Let $\mathbf{y}$ be any additive inverse of $\mathbf{w}$ in S, meaning $\mathbf{y} \in S$ and $\mathbf{w} +_S \mathbf{y} = 0_S$. (There exists at least one such $\mathbf{y}$, by Definition 3.1 applied to S.) $S \subseteq R$ implies $\mathbf{y} \in R$. From Theorem 3.3, $0_S = 0_R$, and $\mathbf{w} +_S \mathbf{y} = \mathbf{w} +_R \mathbf{y}$ by Definition 3.2, so $\mathbf{w} +_R \mathbf{y} = 0_R$, which means $\mathbf{y}$ is an additive inverse of $\mathbf{w}$ in R.

Second, we show that an additive inverse of $\mathbf{w}$ in R is also an additive inverse of $\mathbf{w}$ in S. Let $\mathbf{z}$ be any additive inverse of $\mathbf{w}$ in R, meaning $z \in R$ and $\mathbf{w} +_R \mathbf{z} = 0_R$. (There exists at least one such $\mathbf{z}$, by Definition 3.1 applied to R.) Then $\mathbf{w} +_R \mathbf{z} = 0_R = \mathbf{w} +_R \mathbf{y}$, so by Left Cancellation in R ([BB3], [BB4] Prop. 3.1.7), $\mathbf{z} = \mathbf{y}$ and $\mathbf{y} \in S$, which imply $\mathbf{z} \in S$ and $\mathbf{w} +_S \mathbf{z} = \mathbf{w} +_S \mathbf{y} = 0_S$, meaning $\mathbf{z}$ is an additive inverse of $\mathbf{w}$ in S.

By uniqueness of opposites (Exercise 2.11 applied to either R or S), we can refer to $\mathbf{y} = \mathbf{z}$ as “the” opposite of $\mathbf{w}$, and denote it $\mathbf{y} = -\mathbf{w}$. ■

Theorem 3.4 also implies that subtraction in S is the same as subtraction in R: by the above definition, for $\mathbf{v}$, $\mathbf{w} \in S$, $\mathbf{v} -_S \mathbf{w} = \mathbf{v} +_S \mathbf{y} = \mathbf{v} +_R \mathbf{y} = \mathbf{v} -_R \mathbf{w}$.

Theorem 3.4 can be used in this way: if S is a subset of a rng R and there is an element $\mathbf{w} \in S$, where the opposite of $\mathbf{w}$ in R is not an element of S, then S is not a subrng of R.

Theorem 3.5. *Let $(R, +_R, \cdot_R, 0_R)$ be a rng, and let S be a subset of R. Then S, with the same addition and multiplication operations, is a subrng of R if and only if:*

(1) $\mathbf{x} \in S$, $\mathbf{y} \in S$ imply $\mathbf{x} +_R \mathbf{y} \in S$ (closure under $+_R$ addition), and

(2) $\mathbf{x} \in S$ implies the opposite of $\mathbf{x}$ in R, $-_R\mathbf{x}$, is an element of S (closure under $-_R$ opposite), and

(3) $\mathbf{x} \in S$, $\mathbf{y} \in S$ imply $\mathbf{x} \cdot_R \mathbf{y} \in S$ (closure under $\cdot_R$ multiplication), and

(4) $S \neq \emptyset$.

Proof. Let R have zero element 0_R.

First suppose S is a subrng, so that as in the Proof of Theorem 3.3, S contains a zero element 0_S, which shows $S \neq \emptyset$, and (4) is true. From Definition 3.1, $\mathbf{x} \in S$, $\mathbf{y} \in S$ imply $\mathbf{x} +_S \mathbf{y} \in S$, and from Definition 3.2, $\mathbf{x} +_S \mathbf{y} = \mathbf{x} +_R \mathbf{y}$, so $\mathbf{x} +_R \mathbf{y} \in S$, establishing (1). From Definition 3.1, $\mathbf{x} \in S$ implies $-_S\mathbf{x} \in S$, and by Theorem 3.4, because S is a subrng, $-_S\mathbf{x} = -_R\mathbf{x}$, so $-_R\mathbf{x} \in S$, establishing (2). Similarly, from Definition 3.1, $\mathbf{x} \in S$, $\mathbf{y} \in S$ implies $\mathbf{x} \cdot_S \mathbf{y} \in S$, and from the definition of subrng, $\mathbf{x} \cdot_S \mathbf{y} = \mathbf{x} \cdot_R \mathbf{y}$, so $\mathbf{x} \cdot_R \mathbf{y} \in S$, establishing (3).

Conversely, it follows from (1), (2), (3), and (4) that S is a subrng of R, as follows: R is a rng, and S is a subset of R by hypothesis. Define $+_S$ and $\cdot_S$ by $\mathbf{x} +_S \mathbf{y} = \mathbf{x} +_R \mathbf{y}$, and $\mathbf{x} \cdot_S \mathbf{y} = \mathbf{x} \cdot_R \mathbf{y}$ — these define operations on S by (1) and (2) (so S is closed under $+_S$ and $\cdot_S$, which is required for S to be a rng as in Definitions 3.1 and 3.2), but it remains to check the other properties to show that $(S, +_S, \cdot_S)$ is a rng. Since $S \neq \emptyset$ by (4), there is some $\mathbf{x} \in S$, and by (2), $-_R\mathbf{x} \in S$. By (1), $\mathbf{x} +_R (-_R\mathbf{x}) = 0_R \in S$. $0_R \in S$ satisfies $\mathbf{x} +_S 0_R = \mathbf{x} +_R 0_R = \mathbf{x}$ for all $\mathbf{x} \in S$, so 0_R is a zero element for S. It follows that $-_R\mathbf{x}$ is an additive inverse of $\mathbf{x}$ in S: $\mathbf{x} +_S (-_R\mathbf{x}) = \mathbf{x} +_R (-_R\mathbf{x}) = 0_R = 0_S$. The other rng properties, associativity of $+_S$ and $\cdot_S$, commutativity of $+_S$, and the distributive properties, follow immediately from the facts that these properties hold in R and the operations in S give the same sums and products. ■

Property (2) in Theorem 3.5 is necessary; for example, the set of even integers is a rng, and the subset of positive evens is non-empty and closed under addition and multiplication, but is not a subrng.

Definition 3.6. Given a rng R and a subset $I \subseteq R$, I is an ideal subrng means: I is a subrng, with the property that for any $x \in I$ and any $r \in R$, the products $x \cdot r$ and $r \cdot x$ are in I.

Such a set is also called just an ideal, or a two-sided ideal. If R is a commutative rng, then $x \cdot r \in I$ is enough to imply the other condition. However, when R is not commutative, one can define left ideals requiring only $r \cdot x \in I$, and right ideals where $x \cdot r \in I$.

To tell whether a set I is an ideal of R, it's enough to check some of the properties, and then the rest follow automatically (as in Theorem 3.5):

Proposition 3.7. *Given a rng R and a subset $I \subseteq R$, I is an ideal of R if and only if (1) $I \neq \emptyset$, (2) if $x, y \in I$, then $x +_R y \in I$, (3) if $a \in I$, then its additive inverse $-_R a$ in R satisfies $-_R a \in I$, and (4) if $x \in I$ and $r \in R$, then $x \cdot r \in I$ and $r \cdot x \in I$.* ■

Theorem 3.8. *Given any set Y, and any rng R, the set of functions $M = \{f : Y \to R\}$ is a rng. If $B \subseteq Y$ is any subset, and $K \subseteq M$ is a subrng, then the set*

$$\mathcal{I}(B) = \{f \in K : f(x) = 0_R \text{ for all } x \in B\}$$

is an ideal in K. If $B \subseteq C \subseteq Y$, then $\mathcal{I}(C) \subseteq \mathcal{I}(B)$.

Proof. First, the usual sum and product of functions define operations so that M is a rng. By Theorem 3.5, every subrng K of M contains 0_M, which is the constant zero function ($x \mapsto 0_R$ for all $x \in Y$). The set $\mathcal{I}(B)$ always contains 0_M. To show $\mathcal{I}(B)$ is an ideal using Proposition 3.7, it is easy to check that if f and g are in $\mathcal{I}(B)$ then $-f$ and $f + g$ are in $\mathcal{I}(B)$, and if $r \in K$, then $f \cdot r \in \mathcal{I}(B)$.

If $f \in \mathcal{I}(C)$, then $f(x) = 0_R$ for all $x \in C$. In particular, $f(x) = 0_R$ for all $x \in B \subseteq C$, so $f \in \mathcal{I}(B)$. This shows $\mathcal{I}(C) \subseteq \mathcal{I}(B)$. ■

Notation 3.9. Given a rng R, let S be any subset of R. Define $\langle S \rangle \subseteq R$ to be the intersection of all ideals I such that $S \subseteq I$. S is called a generating set or a basis of $\langle S \rangle$. By the following Lemma, $\langle S \rangle$ can be called the ideal generated by S.

Lemma 3.10. *For any set S, the set $\langle S \rangle$ is an ideal in R. If I is an ideal and $S \subseteq I$, then $\langle S \rangle \subseteq I$.*

Sketch of Proof. It is easy to check that the intersection of two ideals is an ideal, and the proof that the intersection of any collection of ideals is an ideal is a straightforward generalization. The second part of the Lemma follows immediately from the definition of intersection, and it means that $\langle S \rangle$ is the "smallest" ideal containing S. ■

Exercise 3.11. Given a rng R, let I and J be ideals in R. Show by an example that $I \cup J$ is not necessarily an ideal in R. Show that $I \cup J \subseteq I + J$. Show that $\langle I \cup J \rangle = I + J$. ■

Definition 3.12. For an ideal I in R, I is a principal ideal means that there exists some $a \in R$ such that $I = \langle \{a\} \rangle$.

Definition 3.13. A rng R is a principal ideal rng means that every ideal I in R is a principal ideal.

Definition 3.14. For an ideal I in R, I is a finitely generated ideal means that there exists some finite set $\{a_1, a_2, \ldots, a_n\} \subseteq R$ such that $I = \langle \{a_1, a_2, \ldots, a_n\} \rangle$.

Notation 3.15. The ideal $\langle \{a_1, a_2, \ldots, a_n\} \rangle$ can be abbreviated $\langle a_1, a_2, \ldots, a_n \rangle$.

Proposition 3.16 ([H], §III.2)**.**

$$\langle a \rangle = \{r \cdot a + a \cdot s + n \cdot a + \sum_{i=1}^{m} r_i \cdot a \cdot s_i : r, s, r_i, s_i \in R, n \in \mathbb{Z}, m \in \mathbb{N}\}.$$

■

As in Notation 2.16, the term $n \cdot a$ refers to $a + a + \ldots + a$ (n terms) if $n \in \mathbb{N}$, $0 \in R$ if $n = 0$, and $(-a) + (-a) + \ldots + (-a)$ ($-n$ terms) if $-n \in \mathbb{N}$.

Proposition 3.17. *Given a rng R, if a is an element of the center of R (meaning $a \cdot r = r \cdot a$ for all $r \in R$) then*

$$\langle a \rangle = \{r \cdot a + n \cdot a : r \in R, n \in \mathbb{Z}\}.$$

∎

For example, in a commutative rng R, every $a \in R$ is in the center.

Proposition 3.18. *Given a rng R and an element $a \in R$, the set $Ra = \{r \cdot a : r \in R\}$ is a left ideal of R.* ∎

However, the set Ra need not contain a, for example if $R = 2\mathbb{Z}$ and $a = 2$, then $a \notin Ra = 4\mathbb{Z}$.

Theorem 3.19. *Given a rng R, let I and J be ideals in R. The following set is an ideal:*

$$\left\{ \sum_{i=1}^{n} f_i g_i : f_i \in I, g_i \in J, n \in \mathbb{N} \right\},$$

and it is equal to the ideal $\langle \{f \cdot g : f \in I, g \in J\} \rangle$.

Proof. To show it's an ideal, use Proposition 3.7. The set contains the zero element: $0 = 0 \cdot 0$, a sum with one term ($n = 1$). The set is closed under sums: for any $a_1, \ldots a_k \in I$, $b_1, \ldots b_k \in J$, $A_1, \ldots, A_j \in I$, $B_1, \ldots B_j \in J$, define lists α_i and β_i by $\alpha_i = a_i$ if $1 \le i \le k$, $\alpha_i = A_{i-k}$ if $k + 1 \le i \le k + j$, and $\beta_i = b_i$ if $1 \le i \le k$, $\beta_i = B_{i-k}$ if $k + 1 \le i \le k + j$. Then $\alpha_i \in I$, and $\beta_i \in J$, and

$$\sum_{\ell=1}^{k} a_\ell b_\ell + \sum_{m=1}^{j} A_m B_m = \sum_{i=1}^{k+j} \alpha_i \beta_i.$$

The set is also closed under additive inverses: $-\sum a_i b_i = \sum(-a_i) b_i$, since $-a_i \in I$. Finally, it has the two-sided ideal properties: for any $r \in R$, $(\sum a_i b_i) \cdot r = \sum a_i (b_i \cdot r)$ is in the set, since $b_i \cdot r \in J$, and similarly, $r \cdot (\sum a_i b_i) = \sum (r \cdot a_i) b_i$ is in the set, since $r \cdot a_i \in I$.

So, this set is an ideal, and it contains every product of the form $f \cdot g$ (single-term sums), and Lemma 3.10 says $\langle \{f \cdot g\} \rangle \subseteq \{\sum f_i g_i\}$. To show the other subset relation, consider any $\sum f_i g_i$, with $f_i \in I$, $g_i \in J$. If X is any ideal containing all the products $\{f \cdot g\}$, then $f_i \cdot g_i \in X$, and X is closed under sums, so $\sum f_i g_i \in X$. So, any ideal containing all the products $\{f \cdot g\}$ must contain the sum $\sum f_i g_i$, and the intersection of all of these ideals X must also contain $\sum f_i g_i$, which means $\sum f_i g_i \in \langle \{f \cdot g : f \in I, g \in J\} \rangle$. This shows the sets are equal. ∎

Definition 3.20. The ideal from the previous Theorem is denoted IJ, the product ideal.

Exercise 3.21. Given a rng R, and I, J ideals in R, $IJ \subseteq I \cap J$. ∎

Definition 3.22. Given a commutative rng R and any ideal I in R, the radical of I is the set

$$\sqrt{I} = \{f \in R : \exists m \in \mathbb{N} : f^m \in I\}.$$

The quantity f^m refers to f multiplied by itself m times, as in Notation 2.16. Note $I \subseteq \sqrt{I} \subseteq R$.

Exercise 3.23. Given a commutative rng R and any ideal I in R, $\sqrt{I}$ is an ideal in R. ■

Definition 3.24. Given a commutative rng R and any ideal I in R, I is a radical ideal means that $I = \sqrt{I}$.

Exercise 3.25. Given a commutative rng R and any ideal I in R, $\sqrt{\sqrt{I}} = \sqrt{I}$, so $\sqrt{I}$ is always a radical ideal. ■

Exercise 3.26. Given a commutative rng R and any ideals I, J in R, show that $\sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}$. ■

4 Rings

Definition 4.1. A set S, together with two binary operations $+_S$ and $\cdot_S$, is a ring means that $+_S$ and $\cdot_S$ are associative, there is an element $0_S \in S$ which is an identity element for $(S, +_S)$, there is an element $1_S \in S$ which is an identity element for $(S, \cdot_S)$, and every $x \in S$ has a $+_S$-inverse, denoted $-_S x$, or just $-x$. The two operations (abbreviated $+$ and $\cdot$) are related by the distributive laws: for all $a, b, c \in S$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

Theorem 4.2. *If S is a ring, then S is a rng.*

Proof. The only part of the definition of rng missing from Definition 4.1 is the commutativity of $+_S$. Let 1 be the identity element for $\cdot_S$, and consider $\mathbf{v}, \mathbf{w} \in S$. We start with $(1+1)\cdot(\mathbf{v}+\mathbf{w}) \in S$, set LHS=RHS, and use both distributive laws:

$$\begin{aligned}
(1+1)\cdot(\mathbf{v}+\mathbf{w}) &= (1+1)\cdot(\mathbf{v}+\mathbf{w}) \\
((1+1)\cdot\mathbf{v}) + ((1+1)\cdot\mathbf{w}) &= (1\cdot(\mathbf{v}+\mathbf{w})) + (1\cdot(\mathbf{v}+\mathbf{w})) \\
((1\cdot\mathbf{v}) + (1\cdot\mathbf{v})) + ((1\cdot\mathbf{w}) + (1\cdot\mathbf{w})) &= (\mathbf{v}+\mathbf{w}) + (\mathbf{v}+\mathbf{w}) \\
(\mathbf{v}+\mathbf{v}) + (\mathbf{w}+\mathbf{w}) &= (\mathbf{v}+\mathbf{w}) + (\mathbf{v}+\mathbf{w}).
\end{aligned}$$

Then, the associative law gives $\mathbf{v} + (\mathbf{v} + (\mathbf{w} + \mathbf{w})) = \mathbf{v} + (\mathbf{w} + (\mathbf{v} + \mathbf{w}))$, and Left Cancellation leaves $\mathbf{v} + (\mathbf{w} + \mathbf{w}) = \mathbf{w} + (\mathbf{v} + \mathbf{w})$. Using the associative law again, $(\mathbf{v} + \mathbf{w}) + \mathbf{w} = (\mathbf{w} + \mathbf{v}) + \mathbf{w}$, and Right Cancellation gives the result $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$. ■

So, the addition operation for any ring S defines an abelian group $(S, +, 0)$. A ring S where, additionally, $\cdot_S$ is a commutative operation is called a commutative ring.

Some authors use the term "ring" to mean a rng as in Definition 3.1, and then use the above Definition 4.1 to define a "ring with identity."

Definition 4.3. Let $(R, +_R, \cdot_R, 0_R, 1_R)$ be a ring. A set S is a subring of R means:

- $S \subseteq R$, and
- There are operations $+_S$ and $\cdot_S$ such that $(S, +_S, \cdot_S, 0_S, 1_S)$ is a ring, and
- For all $\mathbf{x}$, $\mathbf{y} \in S$, $\mathbf{x} +_R \mathbf{y} = \mathbf{x} +_S \mathbf{y}$, and
- For all $\mathbf{x}$, $\mathbf{y} \in S$, $\mathbf{x} \cdot_R \mathbf{y} = \mathbf{x} \cdot_S \mathbf{y}$, and

- $1_S = 1_R$.

Note that it follows from Definition 4.3 that if S is a subring of R, then for x, $y \in S$, the sum and product $x +_R y$ and $x \cdot_R y$ are in S. The converse of this statement is false (for example, the positive integers $\mathbb{N}$ are closed under sums and products in $\mathbb{Z}$, but $\mathbb{N}$ isn't a ring). The following Theorem claims that closure under sums, products, and additive inverses, and containment of R's multiplicative identity element, is enough for non-empty subsets to be subrings.

Theorem 4.4. *Let $(R, +_R, \cdot_R, 0_R, 1_R)$ be a ring, and let S be a subset of R. The following are equivalent:*

- *S is a subring of R;*
- *S is a subrng of R and $1_R \in S$;*
- *(1) $\mathbf{x} \in S$, $\mathbf{y} \in S$ imply $\mathbf{x} +_R \mathbf{y} \in S$ (closure under $+_R$ addition), and (2) $\mathbf{x} \in S$ implies the opposite of $\mathbf{x}$ in R, $-_R\mathbf{x}$, is an element of S (closure under $-_R$ opposite), and (3) $\mathbf{x} \in S$, $\mathbf{y} \in S$ imply $\mathbf{x} \cdot_R \mathbf{y} \in S$ (closure under $\cdot_R$ multiplication), and (4) $1_R \in S$.*

Proof. Assuming $1_R \in S$, it follows that S is non-empty, and the equivalence of the second and third statements is Theorem 3.5.

Suppose S is a subring of R. Then S is a ring by Definition 4.3, and R and S are rngs by Theorem 4.2, so S is a subrng of the rng R as in Definition 3.2. S contains 1_S because S is a ring, $1_S = 1_R$ by Definition 4.3, so $1_R \in S$.

Conversely, suppose S is a subrng of R and $1_R \in S$. $0_S = 0_R$ as in Theorem 3.5. For all $s \in S$, $1_R \cdot_S s = 1_R \cdot_R s = s = s \cdot_R 1_R = s \cdot_S 1_R$, so 1_R is an identity element for $\cdot_S$: let $1_S = 1_R$. So, S is a rng with a multiplicative identity: $(S, +_S, \cdot_S, 0_R, 1_R)$ is a ring, and satisfies the definition of subring. ∎

For the rest of this Section, let R be a ring, with operations $+$ and $\cdot$, additive identity $0_R = 0$, multiplicative identity $1_R = 1$, and additive inverses $-x$.

Exercise 4.5. Given a ring R, and an ideal I in R, the following are equivalent.

- $1 \in I$.
- I is a subring of R.
- $I = R$.

∎

Proposition 4.6. *Given a ring R and an element $a \in R$,*

$$\langle a \rangle = \{\sum_{i=1}^{m} r_i \cdot a \cdot s_i : r_i, s_i \in R, m \in \mathbb{N}\}.$$

∎

Proposition 4.7. *Given a ring R, and an element $a \in R$, the set $Ra = \{r \cdot a : r \in R\}$ is a left ideal of R, and $a \in Ra$.*

Proof. This follows from Proposition 3.18, and because $1 \in R$, $1 \cdot a = a \in Ra$. ∎

Proposition 4.8. *Given a ring R and an element a of the center of R, then $\langle a \rangle = Ra = aR$.* ∎

Proposition 4.9. *Given a ring R and any non-empty subset X in the center of R, then*

$$\langle X \rangle = \{\sum_{i=1}^{n} r_i \cdot x_i : r_i \in R, x_i \in X, n \in \mathbb{N}\}.$$

∎

Note that X could be an infinite set, but all the elements of $\langle X \rangle$ are just finite sums. As a trivial case, $\langle\ \emptyset\ \rangle = \{0_R\}$.

Definition 4.10. A commutative ring R is an <u>integral domain</u> means that for all $a, b \in R$, $ab = 0 \implies a = 0$ or $b = 0$.

Lemma 4.11. *If R is an integral domain, then $\{0_R\}$ is a radical ideal.*

Proof. Suppose R is an integral domain. Then, by Definition 3.24, "$\{0_R\}$ is a radical ideal" means that if $a^n \in \{0_R\}$, then $a \in \{0_R\}$. Elements a such that $a^n = 0_R$ are called <u>nilpotent</u>, and we need to show that 0_R is the only nilpotent element of R. Statements involving $n \in \mathbb{N}$ require induction: if $n = 1$, then $a^1 = 0_R \implies a = 0_R$, so the $n = 1$ case is true. Suppose that the only element $x \in R$ such that $x^n = 0_R$ is $x = 0_R$. If $a^{n+1} = 0_R$, then $a \cdot a^n = 0_R$, and by the integral domain property, either $a = 0_R$, or $a^n = 0_R$. Either way, $a = 0_R$ by the inductive hypothesis. ∎

Theorem 4.12. *Given any set Y, and any ring R, the set of functions $M = \{f : Y \to R\}$ is a ring. If $B \subseteq Y$ is any subset, and $K \subseteq M$ is a subring, then the set $\mathcal{I}(B) = \{f \in K : f(x) = 0_R$ for all $x \in B\}$ is an ideal in K. If $B \subseteq C \subseteq Y$, then $\mathcal{I}(C) \subseteq \mathcal{I}(B)$. If R is an integral domain, then $\mathcal{I}(B)$ is a radical ideal.*

Proof. First, the usual sum and product of functions define operations so that M is a ring, as in Theorem 3.8. By Definition 4.3, every subring K of M contains 1_M, which is the constant function ($x \mapsto 1_R$ for all $x \in Y$). The first two claims about the ideal $\mathcal{I}(B)$ are exactly as in Theorem 3.8.

To show that $\mathcal{I}(B)$ is radical, suppose $f \in \sqrt{\mathcal{I}(B)}$, so that $f^m \in \mathcal{I}(B)$. Then, for any $x \in B$, $f^m(x) = 0_R = (f(x))^m$, and by Lemma 4.11 about integral domains, $f(x) = 0_R$, so $f \in \mathcal{I}(B)$. ∎

Definition 4.13. A ring R is a <u>principal ideal domain</u> (also called a <u>p.i.d.</u>) means that R is an integral domain and is also a principal ideal rng.

Definition 4.14. An ideal I of a commutative ring R is a <u>prime</u> ideal means: (1) $I \neq R$, and (2) if $a, b \in R$ and $a \cdot b \in I$, then $a \in I$ or $b \in I$.

Example 4.15. If $p \in \mathbb{Z}$ is a "prime number," then $p\mathbb{Z} \subseteq \mathbb{Z}$ is a prime ideal.

Proposition 4.16. *If R is a commutative ring and $R \neq \{0_R\}$, then the following are equivalent: R is an integral domain $\iff$ $\{0_R\}$ is a prime ideal.* $\blacksquare$

(Compare this with Lemma 4.11.)

Exercise 4.17. If R is a commutative ring, and I is a prime ideal, then I is a radical ideal. $\blacksquare$

Exercise 4.18. Given a commutative ring R, if I is a prime ideal and there are N elements $a_1, \ldots, a_N \in R$ so that $a_1 \cdot \ldots \cdot a_N \in I$, then at least one of the elements a_k is in I. $\blacksquare$

Exercise 4.19. Prove that every prime ideal I in R contains all the nilpotent elements of R.

Hint. This means if $a^n = 0_R$, then $a \in I$. One could prove this statement by induction on n, as in Lemma 4.11. $\blacksquare$

Definition 4.20. An ideal I of the commutative ring R is a <u>maximal</u> ideal means: (1) $I \neq R$, and (2) if J is an ideal of R and $I \subseteq J \subseteq R$, then $J = I$ or $J = R$.

Proposition 4.21. *If R is a commutative ring, and I is a maximal ideal, then I is a prime ideal.* $\blacksquare$

Proposition 4.22. *If R is a p.i.d. and I is a prime ideal of R, then either $I = \{0_R\}$, or I is a maximal ideal.* $\blacksquare$

Definition 4.23. A commutative ring R is a <u>field</u> means: $0_R \neq 1_R$, and for any element $s \neq 0_R$ there is an element $r \in R$ such that $s \cdot_R r = 1_R$.

Exercise 4.24. If R is a field then R is an integral domain. $\blacksquare$

5 Very elementary algebraic geometry

Notation 5.1. For a field F with additive identity 0 and multiplicative identity 1, let F^n denote the set of ordered n-tuples of elements of F. Let $F[x_1, \ldots, x_n] = F[\vec{x}]$ denote the set of polynomials in variables $x_1, \ldots, x_n$ with coefficients in F.

F^n is called an <u>n-dimensional affine space</u>. It's also a vector space, but we won't need that extra information. Note that $F[\vec{x}]$ is a commutative ring (it contains the constant polynomials 0 and 1).

Proposition 5.2. *$F[x_1]$, the polynomial ring in one variable, is a p.i.d.* $\blacksquare$

Proposition 5.3. *For $n > 1$, $F[x_1, \ldots, x_n]$ is an integral domain, but not a p.i.d.* $\blacksquare$

Exercise 5.4. $\mathbb{Z}[x_1]$ is an integral domain, but not a p.i.d. $\blacksquare$

Proposition 5.5. *$F[x_1, \ldots, x_n]$ is a unique factorization domain.* $\blacksquare$

This means that any polynomial f can be factored as a product of irreducible polynomials, which are polynomials that cannot be factored as products of non-constant polynomials, and that for $f \neq 0$, the factorization is unique up to re-ordering and scalar multiplication.

Theorem 5.6. *For $f = f(x_1, \dots, x_n) \in F[\vec{x}]$, the following are equivalent: f is a non-constant irreducible polynomial $\iff$ $\langle f \rangle$ is a prime ideal in $F[\vec{x}]$.*

Proof. Note f non-constant is equivalent to $\langle f \rangle \subsetneq F[\vec{x}]$.

First, assuming f is irreducible, suppose a and b are polynomials and $a \cdot b \in \langle f \rangle$. By Proposition 4.8, every element of $\langle f \rangle$ is a multiple of f, and by Proposition 5.5, the following factorization into irreducible polynomials is unique:

$$a \cdot b = m \cdot f = (a_1 \cdots a_k) \cdot (b_1 \cdots b_\ell) = m_1 \cdots m_{k+\ell-1} \cdot f.$$

So one of the LHS factors $a_1, \dots, b_\ell$ is a scalar multiple of f and either a or b is an element of $\langle f \rangle$.

Conversely, suppose $\langle f \rangle$ is a prime ideal and $f = a \cdot b$ for some polynomials a and b. Then $a \cdot b \in \langle f \rangle$, so either $a \in \langle f \rangle$ or $b \in \langle f \rangle$. In the first case, $a = f \cdot c$ for some c by Proposition 4.8, so $f = (f \cdot c) \cdot b$ and so b is a constant, by unique factorization. Similarly a is constant in the second case, so f is an irreducible polynomial. ∎

Proposition 5.7 (Hilbert Basis Theorem)**.** *Every ideal J in $F[\vec{x}]$ is finitely generated.* ∎

Definition 5.8. For any ideal J in $F[\vec{x}]$, define

$$\mathcal{Z}(J) = \{\vec{x} \in F^n : f(\vec{x}) = 0 \text{ for all } f \in J\}.$$

Definition 5.9. A subset $V \subseteq F^n$ is an algebraic set means that there are finitely many polynomials $f_1(\vec{x}), \dots, f_N(\vec{x}) \in F[\vec{x}]$ so that

$$V = \{\vec{x} \in F^n : f_1(\vec{x}) = \cdots = f_N(\vec{x}) = 0\}. \tag{1}$$

V is also called a "variety," an "algebraic variety," or an "affine variety," or the "solution set" or "zero set" of the system of polynomial equations $f_1 = \cdots = f_N = 0$.

Lemma 5.10. *Any algebraic set V is of the form $V = \mathcal{Z}(J)$ for some ideal J in $F[\vec{x}]$.*

Proof. Given V as in (1), let J be the ideal $J = \langle f_1, \dots, f_N \rangle$ as in Notation 3.15. If $\vec{x} \in \mathcal{Z}(J)$, then $f_1(\vec{x}) = \ldots = f_N(\vec{x}) = 0$ since $f_1, \dots, f_N$ are all elements of J, and so $x \in V$ by definition of V, and $\mathcal{Z}(J) \subseteq V$. Conversely, if $\vec{x} \in V$, then $f_1(\vec{x}) = \ldots = f_N(\vec{x}) = 0$, and by Proposition 4.9, any element $f \in J$ is of the form

$$f = \sum_{i=1}^{N} f_i \cdot r_i,$$

so $f(\vec{x}) = \sum 0 \cdot r_i(\vec{x}) = 0$, and $\vec{x} \in \mathcal{Z}(J)$. This shows $V \subseteq \mathcal{Z}(J)$ and $V = \mathcal{Z}(J)$. As special cases, $F^n = \mathcal{Z}(\langle 0 \rangle)$ and $\emptyset = \mathcal{Z}(F[\vec{x}]) = \mathcal{Z}(\langle 1 \rangle)$. ∎

Theorem 5.11. *For any ideal J in $F[\vec{x}]$, $\mathcal{Z}(J)$ is an algebraic set.*

Proof. By the Basis Theorem (Proposition 5.7), there exist finitely many polynomials $f_1, \dots, f_N$ so that $J = \langle f_1, \dots, f_N \rangle$. Let V denote the algebraic set defined by these N polynomials, and then $\mathcal{Z}(J) = V$ by the Proof of Lemma 5.10. ∎

Notation 5.12. For an ideal J generated by one polynomial f as in Definition 3.12, in analogy with Notation 3.15, $\mathcal{Z}(J)=\mathcal{Z}(\langle\{f\}\rangle)=\mathcal{Z}(\langle f\rangle)$ can be abbreviated $\mathcal{Z}(f)$, and as in Theorem 5.11, $\mathcal{Z}(f)$ is exactly the zero set of the polynomial f.

Theorem 5.13. *If J_1 and J_2 are ideals in $F[\vec{x}]$ and $J_1 \subseteq J_2$, then $\mathcal{Z}(J_2)\subseteq\mathcal{Z}(J_1)$.*

Proof. Suppose $\vec{x}\in\mathcal{Z}(J_2)$, which means that $f(\vec{x})=0$ for all $f\in J_2$. Then, since $J_1\subseteq J_2$, $f(\vec{x})=0$ for all $f\in J_1$, and $\vec{x}\in\mathcal{Z}(J_1)$. ∎

So, the "subset" relation is reversed by the "operation" of forming an algebraic set. The idea is that J_1 contains fewer polynomials, and its solution set will be larger than the solution set of a system with more polynomial equations that must be satisfied.

Corollary 5.14. *If $J_1=J_2$, then $\mathcal{Z}(J_1)=\mathcal{Z}(J_2)$.*

Proof. This is obvious (from using the previous Theorem on $J_1\subseteq J_2$ and $J_2\subseteq J_1$), but geometrically it means that if $\{f_1,\ldots,f_N\}$ and $\{g_1,\ldots,g_M\}$ generate the same ideal, then they have the same solution set. ∎

The converse is false: two different ideals can define the same algebraic set.

Exercise 5.15. The intersection of two algebraic sets is an algebraic set.

Hint. This is related to Exercise 3.11. ∎

Theorem 5.16. *If I and J are ideals in $F[\vec{x}]$, then $\mathcal{Z}(IJ)=(\mathcal{Z}(I))\cup(\mathcal{Z}(J))$.*

Proof. First, to show $\subseteq$, suppose $\vec{x}\in\mathcal{Z}(IJ)$, so that $k(\vec{x})=0$ for all $k\in IJ$, and in particular $f(\vec{x})\cdot g(\vec{x})=0$ for all $f\in I$ and $g\in J$. One of the factors must be zero ($f(\vec{x}),g(\vec{x})\in F$, and every field is an integral domain). If $f(\vec{x})=0$ for all $f\in I$, then $\vec{x}\in\mathcal{Z}(I)$. Otherwise, there's some f so that $f(\vec{x})\neq 0$, in which case $f(\vec{x})\cdot g(\vec{x})=0 \implies g(\vec{x})=0$ for all $g\in J$, so $\vec{x}\in\mathcal{Z}(J)$. Either way, $\vec{x}\in\mathcal{Z}(I)\cup\mathcal{Z}(J)$.

Also, to show the other inclusion, suppose $\vec{x}\in\mathcal{Z}(I)$, so that $f(\vec{x})=0$ for all $f\in I$. Any element in IJ is of the form $h=\sum f\cdot g$, so $h(\vec{x})=\sum f(\vec{x})\cdot g(\vec{x})=\sum 0\cdot g(\vec{x})=0 \implies \vec{x}\in\mathcal{Z}(IJ)$. Similarly, assuming $\vec{x}\in\mathcal{Z}(J)$ will imply $\vec{x}\in\mathcal{Z}(IJ)$. ∎

Exercise 5.17. If I and J are ideals in $F[\vec{x}]$, then $\mathcal{Z}(I\cap J)=(\mathcal{Z}(I))\cup(\mathcal{Z}(J))$. It follows that the union of algebraic sets is an algebraic set.

Hint. Note this is the same union as in the previous Theorem! Use Exercise 3.21, to get $IJ\subseteq I\cap J\subseteq I$, and then apply Theorem 5.13. Also, use Theorem 5.13 with $I\cap J\subseteq J$. ∎

Example 5.18. Find the solution set of the equations

$$\begin{aligned} 3x+5y+2 &= 0\\ 15x+9y+6 &= 0. \end{aligned}$$

These are two linear polynomials, $\{f_1,f_2\}$ in $\mathbb{Q}[x,y]$. One approach is to consider this as a linear algebra problem and rewrite it as a matrix equation:

$$\begin{pmatrix} 3 & 5 & 2\\ 15 & 9 & 6\end{pmatrix}\cdot\begin{pmatrix} x\\ y\\ 1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

Finding the solutions $(x, y) \in \mathbb{Q}^2$ is equivalent to finding the "kernel" of the matrix.

A different approach to the problem is to multiply the first equation by 5 and then subtract it from the second equation, to get the x terms to cancel, leaving only $-16y - 4 = 0$. The solution is $y = -\frac{1}{4}$, and then plugging this into the first equation determines $x = -\frac{1}{4}$, and $(-\frac{1}{4}, -\frac{1}{4})$ is the only solution.

In terms of the matrix problem, the "subtracting equations" is really the "row-reduction" procedure, which corresponds to multiplying the 2×3 coefficient matrix by $\begin{pmatrix} 1 & 0 \\ -5 & 1 \end{pmatrix}$ on its left. This simplifies the matrix, without changing the kernel.

In terms of ideals, $\{f_1, f_2\}$ generate an ideal $J = \langle f_1, f_2 \rangle \subseteq \mathbb{Q}[x, y]$, and $\mathcal{Z}(J)$ is the solution set. In this case, $\mathcal{Z}(J)$ is the one-point set $\{(-\frac{1}{4}, -\frac{1}{4})\} \subseteq \mathbb{Q}^2$. The linear polynomial $g = f_2 - 5f_1$ is an element of J, and in fact $J = \langle f_1, g \rangle$. (Since $f_1 \in J$, and $g \in J$, it follows that $\langle f_1, g \rangle \subseteq J$. Also, f_1 and $f_2 = g + 5f_1$ are in $\langle f_1, g \rangle$, so $J = \langle f_1, f_2 \rangle \subseteq \langle f_1, g \rangle$.) The two approaches to solving the two linear equations, the matrix method, and the subtracting equations method, both started with f_1, f_2, and found a "simpler" basis $\{f_1, g\}$ for the ideal J. Since $\langle f_1, g \rangle$ is the same ideal as $\langle f_1, f_2 \rangle$, these ideals have the same algebraic set, but it's easier to find the solution set when one of the polynomials (g) involves only one variable.

Polynomials $f \in F[\vec{x}]$ define functions $F^n \to F$, and, as in Theorem 4.12, $F[\vec{x}]$ is a subring of the set of all functions $M = \{f : F^n \to F\}$.

Notation 5.19. For the rest of this Section, the symbol $\mathcal{I}(B)$ (introduced in Theorem 3.8) will only be used for subsets $B \subseteq F^n$, and $\mathcal{I}(B)$ will always mean the set of polynomials in $F[\vec{x}]$ which have value 0 at all points in B.

For example, an ideal J in $F[\vec{x}]$ defines a subset $\mathcal{Z}(J) \subseteq F^n$, and Theorem 3.8 says that this subset of F^n defines an ideal $\mathcal{I}(\mathcal{Z}(J))$ in $F[\vec{x}]$.

Theorem 5.20. $J \subseteq \mathcal{I}(\mathcal{Z}(J))$.

Proof. If $f \in J$, and $\vec{x} \in \mathcal{Z}(J)$, then $f(\vec{x}) = 0$ by definition of $\mathcal{Z}(J)$. So, $f(\vec{x}) = 0$ for all $\vec{x} \in \mathcal{Z}(J)$, and this is the definition of $f \in \mathcal{I}(\mathcal{Z}(J))$. ∎

Sometimes, $\mathcal{I}(\mathcal{Z}(J)) = J$, but this is not true in general. Theorem 4.12 says $\mathcal{I}(\mathcal{Z}(J))$ has to be a radical ideal, and not all ideals in $F[\vec{x}]$ are radical. The following Proposition, then, is quite plausible, but it is only stated for $F = \mathbb{C}$, and is not true for $F = \mathbb{R}$ or $\mathbb{Q}$.

Proposition 5.21 (Hilbert's Nullstellensatz)**.** *If J is an ideal in $\mathbb{C}[\vec{x}]$, then*

$$\mathcal{I}(\mathcal{Z}(J)) = \sqrt{J}.$$

∎

Corollary 5.22. *For any set $V \subseteq F^n$, $V \subseteq \mathcal{Z}(\mathcal{I}(V))$, and if V is an algebraic set, then $V = \mathcal{Z}(\mathcal{I}(V))$.*

Proof. Suppose $x \in V$, then for any $f \in \mathcal{I}(V)$, $f(x) = 0$ so $x \in \mathcal{Z}(\mathcal{I}(V))$. Conversely, if V is an algebraic set then by Lemma 5.10, there is some ideal J so that $V = \mathcal{Z}(J)$. By Theorem 5.20, $J \subseteq \mathcal{I}(\mathcal{Z}(J)) = \mathcal{I}(V)$, and then by Theorem 5.13, $\mathcal{Z}(\mathcal{I}(V)) \subseteq \mathcal{Z}(J) = V$. ∎

Theorem 5.23. *For a set $V \subseteq F^n$, and an algebraic set $W \subseteq F^n$, $V \subseteq W \iff \mathcal{I}(W) \subseteq \mathcal{I}(V)$.*

Proof. The implication $V \subseteq W \implies \mathcal{I}(W) \subseteq \mathcal{I}(V)$ is true for any subsets V, W, and was proved in Theorem 3.8. The other implication starts with assuming $\mathcal{I}(W) \subseteq \mathcal{I}(V)$. By definition of algebraic set, there are polynomials $g_1, \ldots, g_M$, so that $W = \{\vec{x} : g_1(\vec{x}) = \cdots = g_M(\vec{x}) = 0\}$. Each g_i satisfies $g_i(\vec{x}) = 0$ for all $\vec{x} \in W$, so $g_i \in \mathcal{I}(W)$, and by hypothesis, $g_i \in \mathcal{I}(V)$. The definition of $g_i \in \mathcal{I}(V)$ is that for any $\vec{v} \in V$, $g_i(\vec{v}) = 0$, so $g_1(\vec{v}) = \cdots = g_M(\vec{v}) = 0$, and this is the definition of $\vec{v} \in W$, which proves $V \subseteq W$. ■

Corollary 5.24. *For two algebraic sets V, $W \subseteq F^n$, $V = W \iff \mathcal{I}(W) = \mathcal{I}(V)$.* ■

Theorem 5.25. *For a point $(a_1, \ldots, a_n) \in F^n$, the ideal*

$$I = \langle x_1 - a_1, \ldots, x_n - a_n \rangle \subseteq F[\vec{x}]$$

is a maximal ideal.

Sketch of Proof. The geometric interpretation is that since the point is a “smallest” non-empty algebraic set, the ideal $\mathcal{I}(\{(a_1, \ldots, a_n)\})$ should be a “largest” ideal strictly contained in $F[\vec{x}]$. There is an elementary but technical proof using just algebra. ■

Proposition 5.26. *If I is a maximal ideal in $\mathbb{C}[\vec{x}]$, then there is some point $(a_1, \ldots, a_n) \in \mathbb{C}^n$ so that*

$$I = \langle x_1 - a_1, \ldots, x_n - a_n \rangle.$$

■

This is a difficult theorem which requires the Nullstellensatz or a similar line of reasoning, so I won’t prove it. It is false if $\mathbb{C}$ is replaced by $\mathbb{R}$ or $\mathbb{Q}$.

Definition 5.27. An algebraic set $V \subseteq F^n$ is irreducible means: if $V = V_1 \cup V_2$ for algebraic sets V_1, V_2, then $V_1 = V$ or $V_2 = V$.

Theorem 5.28. *V is an irreducible algebraic set $\iff \mathcal{I}(V)$ is a prime ideal in $F[\vec{x}]$.*

Proof. First, suppose V is not irreducible, so there are algebraic sets $V_1 \neq V$, $V_2 \neq V$ with $V = V_1 \cup V_2$. By Theorem 3.8, $V_1 \subseteq V \implies \mathcal{I}(V) \subseteq \mathcal{I}(V_1)$. By Corollary 5.24, if $\mathcal{I}(V) = \mathcal{I}(V_1)$ then $V_1 = V$, so $\mathcal{I}(V) \neq \mathcal{I}(V_1)$ and there is some $f \in \mathcal{I}(V_1) \setminus \mathcal{I}(V)$. Similarly, there is some $g \in \mathcal{I}(V_2) \setminus \mathcal{I}(V)$. The product $f \cdot g$ satisfies, for any $\vec{v} \in V$, $(f \cdot g)(\vec{v}) = f(\vec{v}) \cdot g(\vec{v}) = 0$, because either $\vec{v} \in V_1$, so $f(\vec{v}) = 0$, or $\vec{v} \in V_2$, so $g(\vec{v}) = 0$. It follows that $f \cdot g \in \mathcal{I}(V)$, even though neither f nor g is in $\mathcal{I}(V)$, so $\mathcal{I}(V)$ is not a prime ideal.

Second, let $V = \mathcal{Z}(\langle h_1, \ldots, h_N \rangle)$, and suppose $\mathcal{I}(V)$ is not a prime ideal, so there are some $f, g \in F[\vec{x}]$ with $f \cdot g \in \mathcal{I}(V)$ but $f \notin \mathcal{I}(V)$ and $g \notin \mathcal{I}(V)$. Define algebraic sets $V_1 = \mathcal{Z}(\langle f, h_1, \ldots, h_N \rangle)$ and $V_2 = \mathcal{Z}(\langle g, h_1, \ldots, h_N \rangle)$. By Theorem 5.13,

$$\{h_1, \ldots, h_N\} \subseteq \{f, h_1, \ldots, h_N\} \implies \langle \{h_1, \ldots, h_N\} \rangle \subseteq \langle \{f, h_1, \ldots, h_N\} \rangle \implies V_1 \subseteq V.$$

Because $f \notin \mathcal{I}(V)$, there is some $\vec{x} \in V$ with $f(\vec{x}) \neq 0$, so $\vec{x} \notin V_1 = \mathcal{Z}(\langle f, h_1, \ldots, h_N \rangle)$; this shows $V_1 \neq V$. Similarly, $V_2 \subsetneq V$, so $V_1 \cup V_2 \subseteq V$. For any $\vec{v} \in V$, $h_1(\vec{v}) = \ldots = h_N(\vec{v}) = 0$ and because $f \cdot g \in \mathcal{I}(V)$, $(f \cdot g)(\vec{v}) = 0$, so either $f(\vec{v}) = 0 = h_1(\vec{v}) = \ldots = h_N(\vec{v}) \implies \vec{v} \in V_1$, or $g(\vec{v}) = 0 = h_1(\vec{v}) = \ldots = h_N(\vec{v}) \implies \vec{v} \in V_2$. In either case, $\vec{v} \in V_1 \cup V_2$, so $V \subseteq V_1 \cup V_2$ and $V = V_1 \cup V_2$: V is not irreducible. ■

Theorem 5.29. *If J is a prime ideal in $\mathbb{C}[\vec{x}]$ then $\mathcal{Z}(J)$ is an irreducible algebraic set.*

Proof. By Proposition 5.21, $\mathcal{I}(\mathcal{Z}(J)) = \sqrt{J}$, and by Exercise 4.17, because J is a prime ideal, J is also a radical ideal, so $\mathcal{I}(\mathcal{Z}(J)) = \sqrt{J} = J$. It follows that $\mathcal{I}(\mathcal{Z}(J))$ is a prime ideal, so $\mathcal{Z}(J)$ is irreducible by Theorem 5.28. ∎

Example 5.30. Any point in F^n is obviously an irreducible algebraic set. It defines a maximal ideal by Theorem 5.25, and this gives a geometric interpretation of Proposition 4.21.

Example 5.31. Since $F[x_1]$ is a p.i.d. (Proposition 5.2), a geometric interpretation of Proposition 4.22 is that the only irreducible algebraic sets in F^1 are sets with one point, or the empty set, or F^1. Algebraically, it means that a polynomial with more than one root in F is either reducible in $F[x_1]$ (by the Division Theorem), or the constant polynomial 0.

Example 5.32. Let $f(x,y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1 \in \mathbb{R}[x,y]$. Then $f(1,1) = f(1,-1) = f(-1,1) = f(-1,-1) = 0$, and these four points are the only real solutions of $f(x,y) = 0$, so the real algebraic set is a four-point set in the plane $\mathbb{R}^2$:

$$V = \mathcal{Z}(\langle f \rangle) = \{(1,1), (-1,1), (1,-1), (-1,-1)\}.$$

There are no other solutions by the arithmetic-geometric mean inequality:

$$\frac{x^4y^2 + x^2y^4 + 1}{3} \geq \left(x^4y^2 \cdot x^2y^4 \cdot 1\right)^{1/3} = x^2y^2,$$

with equality only if $x^4y^2 = 1$ and $x^2y^4 = 1$.

This V is not irreducible (it is the union $V_1 \cup V_2$ where V_1 is a one point algebraic set, and V_2 is the union of the other three points, which is also an algebraic set).

Comment: this real polynomial f was shown by Motzkin to not be a sum of squares of real polynomials $f_1^2 + f_2^2 + \ldots + f_n^2$, even though its values are always ≥ 0.

References

[BB2] J. Beachy and W. Blair, *Abstract Algebra*, 2nd ed., Waveland Press, Prospect Heights, IL, 1996. `https://faculty.niu.edu/math_beachy/`

[BB3] J. Beachy and W. Blair, *Abstract Algebra*, 3rd ed., Waveland Press, Long Grove, IL, 2006. `https://faculty.niu.edu/math_beachy/`

[BB4] J. Beachy and W. Blair, *Abstract Algebra*, 4th ed., Waveland Press, Long Grove, IL, 2019. `https://faculty.niu.edu/math_beachy/`

[CLO] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*, Undergraduate Texts in Mathematics, Springer, New York, 1992.

[D] J. Durbin, *Modern Algebra. An Introduction.* 3rd ed., John Wiley & Sons, Inc., New York, 1992.

[H] T. Hungerford, *Algebra*, GTM **73**, Springer-Verlag, New York-Berlin, 1980.