Algebraic Geometry

Adam Boocher

Spring 2007

1 Basic Preliminaries

These notes serve as a basic introduction to many of the topics in elementary algebraic geometry. This section starts out rather informally, defining many of the crucial objects in terms of polynomial rings and certain special subsets. We will see however, that hardly anything is lost when we pass to the general case of commutative rings in the next section. The reader should recall the notion of a *polynomial ring* from high school, even if it was not given by that name. All the naive rules concerning addition, subtraction and multiplication of polynomials applies in these notes as it did in high school. The interested reader can of course state these definitions formally.

Definition 1. *The polynomial ring* $\mathbb{R}[x_1, \ldots, x_n]$ *is the set of all polynomials in the* n *variables* $x_1, \ldots, x_n$.

The reason we call this set a ring will be clear in the next section when we define abstract rings, but for the moment it will be helpful to keep in mind that in this set we can do the following:

- add, subtract and multiply polynomials
- the results of these operations are again polynomials
- addition and multiplication satisfy distributive laws

It is important to note that in general we cannot divide two polynomials and get another polynomial. Later, when talking about rings, we will define what it means to be a *unit* and how it is related to dividing elements in a ring, but for now, try this homework problem:

Homework 1. *Let* f, g *be polynomials in* $\mathbb{R}[x_1, \ldots, x_n]$. *Under what conditions is* f/g *a polynomial? For which* g *is* f/g *a polynomial for* all *polynomials* f*? To do this problem it might help to first consider polynomials in one variable (i.e. let* $n = 1$*).*

Polynomial rings are very algebraic objects. This shouldn't come as a surprise since you most likely first heard about them during a high school algebra class. A common theme in Algebraic Geometry is the interplay between algebra and geometry. We now introduce the geometric counterpart of $\mathbb{R}[x_1, \ldots, x_n]$.

Definition 2. *Affine n-space* $\mathbb{A}^n_{\mathbb{R}} = \{(a_1, \ldots, a_n) \ : \ a_i \in \mathbb{R}\} = \mathbb{R}^n$

You'll notice in our definition of $\mathbb{A}^n_{\mathbb{R}}$ we have put a small $\mathbb{R}$ in the subscript. This is because we are using the real numbers as the *field* we are working over. For example, if we were using the complex numbers, then we could easily have said

$$\mathbb{A}^n_{\mathbb{C}} = \mathbb{C}^n.$$

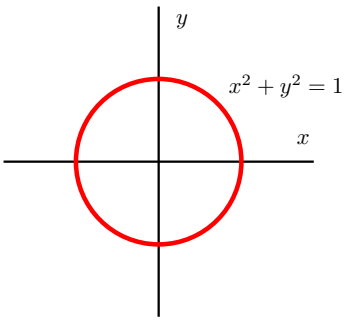
If the underlying field is understood, (or not important), we may omit the subscript and just write $\mathbb{A}^n$. These are mainly technicalities and shouldn't be worried about much. The key is that **affine space is just the standard vector space everyone is used to.** It's perfectly fine to think of it as $\mathbb{R}^n$.

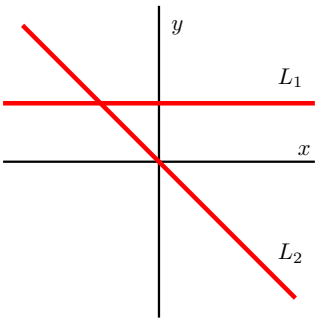
We now begin to discuss the relationship between $\mathbb{R}[x_1, \ldots, x_n]$ and $\mathbb{A}^n_{\mathbb{R}}$. Suppose we have a polynomial, $f \in \mathbb{R}[x_1, \ldots, x_n]$. We define the set

$$V(f) = \{(a_1, \ldots, a_n) \in \mathbb{A}^n_{\mathbb{R}} \ : \ f(a_1, \ldots, a_n) = 0\}$$

This is called the *variety* of f. It is maybe more helpful to think that the V stands for *vanishing*, however. Indeed, $V(f)$ is the set of all points in affine space where f vanishes.

Example: Let us work in $\mathbb{R}[x, y]$. Let $\mathbf{f(x, y) = x^2 + y^2 - 1}$. What is $V(f)$?





Answer: $V(f)$ is the set of all points in $\mathbb{R}^2$ such that $x^2 + y^2 - 1 = 0$. This is easily seen to be the unit circle. Writing it as a set,

$$V(f) = \left\{(x, y) \in \mathbb{R}^2 \ : \ x^2 + y^2 = 1\right\}$$

Example: Staying in $\mathbb{R}[x, y]$. Let $\mathbf{f(x, y) = xy + y^2 - y - x = (x + y)(y - 1)}$. What is $V(f)$?

Answer: $V(f)$ is the set of all points in $\mathbb{R}^2$ such that $(x + y)(y - 1) = 0$. This is the set of all points with $x = -y$ and the set of points with $y = 1$. These are the two lines L_1 and L_2 in the figure. We could write the variety as

$$V(f) = \left\{(x, y) \in \mathbb{R}^2 \ : \ x = -y \ \text{ or } \ y = 1\right\} = L_1 \cup L_2.$$

Homework 2. *What is the difference between these two examples? First consider geometric properties of the varieties. Are there multiple parts? How do these parts correspond to the polynomial? What do the specific components correspond to?*

Homework 3. *In the second example above, every point on the variety is a point where f vanishes. What is special about the point $(-1,-1)$ on the variety? In terms of the polynomial?*

1.1 The other direction:

We just exhibited how any polynomial in $\mathbb{R}[x_1,\ldots,x_n]$ gives rise to a variety in $\mathbb{A}^n_{\mathbb{R}}$. Now let us formulate the question in reverse. Given a subset X of $\mathbb{A}^n_{\mathbb{R}}$. Can we get a polynomial? The answer is yes, in fact we get a whole family of them!

Definition 3. *Let X be a subset of $\mathbb{A}^n_{\mathbb{R}}$. We define the subset $I(X)$ in $\mathbb{R}[x_1,\ldots,x_n]$ to be the set of all polynomials in $\mathbb{R}[x_1,\ldots,x_n]$ that vanish at each point of X. In other words,*

$$I(X) = \{f \in \mathbb{R}[x_1,\ldots,x_n] \ : \ f(a_1,\ldots,a_n) = 0, \text{ for each } (a_1,\ldots,a_n) \in X\}$$

This set has several important properties. We outline them below and will come back to the next section when we discuss *ideals*.

1. $0 \in I(X)$
2. If $f, g \in I(X)$ then so is $f+g$.
3. If $f \in I(x)$ and $h \in \mathbb{R}[x_1,\ldots,x_n]$ is **any** polynomial, then $fh \in I(x)$.

The verification of these three is straightforward, and is left to the reader, but we prove 3, to show the idea.

Proof. Suppose that f and h are as above. Then since $f(a) = 0$ for every $a \in X$, $(fh)(a) = f(a)h(a) = 0h(a) = 0$ for every $a \in X$. Thus $fh \in I(X)$. □

Homework 4. *Prove parts 1 and 2 above. (Note that there is little to prove for part 1, but it will help reinforce the ideas to convince yourself that $0 \in I(X)$).*

Homework 5. *Find a subset of $\mathbb{R}[x_1,\ldots,x_n]$ that satisfies property 2 above, but not property 3. Also, find a subset that satisfies property 3 but not property 2.*

We will call any subset of $\mathbb{R}[x_1,\ldots,x_n]$ that satisfies the above three properties a *polynomial ideal* (or just an *ideal*).

Examples:

- The set $\{0\}$ is an ideal.
- The set (f) of all multiples of f is an ideal. This is the set

$$(f) = \{f(x)g(x) \ : \ g \in \mathbb{R}[x_1,\ldots,x_n]\}$$

- The set $(f,g) = \{f(x)h(x) + g(x)k(x) \ : \ h,k \in \mathbb{R}[x_1,\ldots,x_n]\}$ is an ideal.
- In general we write

$$(f_1,\ldots,f_n) = \{f_1h_1 + \ldots + f_nh_n \ : \ h_i \in \mathbb{R}[x_1,\ldots,x_n]\}$$

Proof. We will prove that (f) is an ideal. Clearly, $0 \in (f)$ since $0 = 0f$. Let $g,h \in (f)$ Then $g = fa, h = fb$ for some polynomials a,b, so

$$g + h = fa + fb = f(a+b) \in (f).$$

Finally, let $g \in (f)$ and h be any polynomial. Then by definition $g = fa$ for some a. Thus

$$gh = fah = f(ah) \in (f).$$

□

Homework 6. *Let $f_1,\ldots,f_n$ be polynomials in $\mathbb{R}[x_1,\ldots,x_n]$. Prove that $(f_1,\ldots,f_n)$ as defined above is an ideal.*

Proposition 1. *Let $f_1,\ldots,f_n$ be polynomials in $\mathbb{R}[x_1,\ldots,x_n]$. Then $(f_1,\ldots,f_n)$ is the smallest ideal containing $f_1,\ldots,f_n$.*

Proof. Let I be any ideal containing $f_1,\ldots,f_n$. We will show that $(f_1,\ldots,f_n) \subseteq I$. Since I contains f_i and is an ideal, it must contain f_ih for every polynomial $h \in \mathbb{R}[x_1,\ldots,x_n]$. Since it is closed under addition as well, it must contain all sums of these guys as well, thus it contains $(f_1,\ldots,f_n)$. □

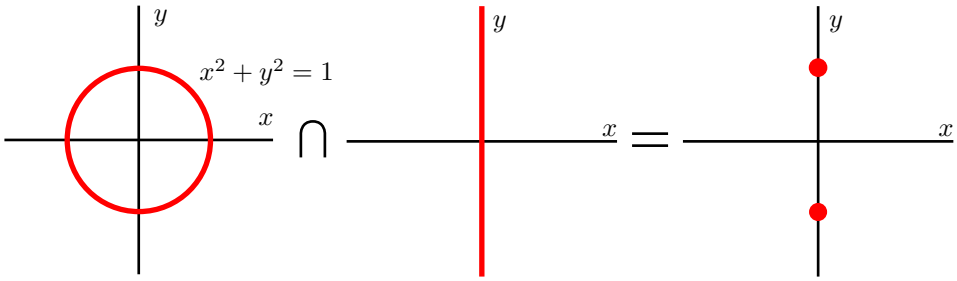
The previous proof used the fact that $(f_1,\ldots,f_n)$ is the set of all "combinations" of the f_i. In fact, it is safe to think about this as a sort of generalized linear combination. The difference from linear algebra is that here we are allowed to multiply by any elements of $\mathbb{R}[x_1,\ldots,x_n]$ so the "coefficients" of the f_i aren't restricted to just constants, but can be polynomials as well.

We now relate the notion of an ideal to the relationship between polynomials and varieties. This is best illustrated in the following proposition:

Proposition 2. *Suppose we define $V(f_1,f_2) = \{a \in \mathbb{A}^n_{\mathbb{R}} \ : \ f_1(a) = 0, f_2(a) = 0\}$. Then this set is the same as $\{a \in \mathbb{A}^n_{\mathbb{R}} \ : \ f(a) = 0$ for each $f \in (f_1,f_2)\}$.*

Remark: The point of this is that if f_1, f_2 vanish at a particular point, then so does every element in the ideal that they generate. So from now on, when we write $V(f)$ there is no fear of confusion because whether we mean the set of points in $\mathbb{A}^n_{\mathbb{R}}$ where f vanishes, or the set of points where *every* polynomial in the ideal (f) vanishes we get the same set.

Proof. Let a be a point of the first type. That is, $f_1(a) = f_2(a) = 0$. Then clearly, all combinations $gf_1 + hf_2$ have $g(a)f_1(a) + h(a)f_2(a) = 0 + 0 = 0$. Conversely, if $f(a) = 0$ for all $f \in (f_1,f_2)$ then certainly $f_1(a) = f_2(a) = 0$. □



Example: Let $f(x, y) = x^2 + y^2 - 1$, $g(x, y) = x$. Then

$$V(f, g) = \{a \in \mathbb{A}^n_{\mathbb{R}} \ : \ f(a) = 0 \ , \ g(a) = 0\} .$$

Unraveling this condition is the same as saying:

$$V(f, g) = \left\{(x, y) \in \mathbb{A}^n_{\mathbb{R}} \ : \ x^2 + y^2 = 1 \ , \ x = 0\right\} .$$

It is then easy to see that $V(f, g) = \{(0, 1), (0, -1)\}$. This solution was largely algebraic so we show the corresponding idea with geometry. The varieties $V(f)$ and $V(g)$ which are the unit circle and the y-axis respectively. It makes sense that the points which vanish on both f and g must be in both of these sets. Thus we take their intersection to obtain $V(f, g)$. This process is illustrated in the figure.

Example: Let $f(x, y) = x^2$. Then $V(f)$ is easily seen to be the y-axis. Now let's go backwards! Let $X = V(f)$ be the y-axis. What is $I(X)$?

$$I(X) = \{f \ : \ f(0, y) = 0 \text{ for each } y\} .$$

Clearly, if $f(x, y) = x \cdot g(x, y)$ then f vanishes on the y-axis and so $f \in I(X)$.

We make the claim that if $I(X)$ consists entirely of these functions. Indeed, suppose that $f \in I(X)$. Then group all terms containing x to the front and write f as

$$f(x, y) = x \cdot p(x, y) + g(y). \ \text{ for some } p, g$$

Then $f(0, y) = 0$ is the same as saying $g(y) = 0$. This equation must be true for all y, which means that g must be the 0 polynomial. Thus $f = x \cdot p$ as required. Thus we have just shown that $I(X) = (x)$, the ideal of all multiples of x. But this implies the following conundrum:

$$I(V(x^2)) = (x).$$

Shouldn't the operations 'I' and 'V' undo each other?

Homework 7. *Prove that if $g(x) = 0$ for infinitely many values of x then g is still the 0 polynomial. Does this latter statement hold true if we allow g to depend on two variables?*

1.2 The Relationship Between I and V

In the previous section we showed a way to switch between ideals of $\mathbb{R}[x_1, \ldots, x_n]$ and subset of $\mathbb{A}^n_{\mathbb{R}}$. Namely we defined functions

$$V : \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\} \longrightarrow \{\text{subsets of } \mathbb{A}^n_{\mathbb{R}}\}$$

$$I : \{\text{subsets of } \mathbb{A}^n_{\mathbb{R}}\} \longrightarrow \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\}$$

Unfortunately, these functions are not bijections, and are certainly not inverses of each other. Indeed, let J be an ideal of $\mathbb{R}[x_1, \ldots, x_n]$. Then as we saw in the previous example, in general,

$$I(V(J)) \neq J.$$

It is not hard to see that in general, if X is a subset of $\mathbb{A}^n_{\mathbb{R}}$, then

$$V(I(X)) \neq X.$$

We illustrate this in the following example.

Example: Let $X = \mathbb{R}^2 - \{(0,0)\}$. Then if $f(x) = 0$ for all $x \in X$ then f is the zero polynomial, so $I(X) = 0$. But then

$$V(I(X)) = V(0) = \mathbb{R}^2.$$

Question 1: What conditions do we need for V and I to be inverses of each other? When is $V(I(X)) = X$ and $I(V(J)) = J$?

Answer: We find that the answers to these questions depend on the type of sets X and the types of ideals J. We will first classify the types of sets X because the development is slightly easier. Although V and I are not inverses of each other, they still behave nicely in for general sets and ideals.

Proposition 3. *1) Let $J \subset \mathbb{R}[x_1, \ldots, x_n]$ be an ideal. Then $J \subset I(V(J))$.*
2) Let $X \subset \mathbb{R}^n$. Then $X \subset V(I(X))$.

Proof. The proofs of these are very simple and just involve translating what the definitions say about I and V. Perhaps more would be gotten out of these proofs if the student wrote them themselves rather than reading them, as reading "by definition" doesn't often give much insight.
1) Let $f \in J$. We'd like to show that $f \in I(V(J))$. To prove this, we'd need to show that $f(x) = 0$ for all $x \in V(J)$. But by definition, if $x \in V(J)$, then $f(x) = 0$ since $f \in J$.
2) Let $x \in X$. Then $f(x) = 0$ for each f in $I(X)$ by definition. Thus $x \in V(I(X))$. □

Proposition 4. *(I and V reverse inclusions)*
1)If $I \subset J$ then $V(I) \supset V(J)$
2)If $X \subset Y$ then $I(X) \supset I(Y)$

Homework 8. *Prove the above proposition.*

The first part of the answer to Question 1 involves defining an *algebraic set*. The following proposition shows that these sets satisfy $V(I(X)) = X$.

Definition 4. *A set $X \subset \mathbb{R}^n$ is called an algebraic set if $X = V(I)$ for some ideal I of $\mathbb{R}[x_1, \ldots, x_n]$.*

Proposition 5. *Let X be an algebraic set. Then $V(I(X)) = X$.*

Remark: Note that if J is an ideal of $\mathbb{R}[x_1, \ldots, x_n]$ then $V(J)$ is an algebraic set by definition. Thus, the proposition states that for all ideals J,

$$V(I(V(J))) = V(J).$$

Proof. Since X is an algebraic set, $X = V(J)$ for some ideal J of $\mathbb{R}[x_1, \ldots, x_n]$. We have already shown that $X \subset V(I(X))$ so what remains is to show that $V(I(X)) \subset X$. Let $x \in V(I(X))$. Then this means that $f(x) = 0$ for every $f \in I(X)$. But $I(X) = I(V(J)) \supset J$, by Proposition 3. Thus since $f(x) = 0$ for every $f \in I(X)$ and $I(X)$ contains J, we certainly have $f(x) = 0$ for every $f \in J$. Thus $x \in V(J) = X$. □

Homework 9. *Prove that not all subsets of $\mathbb{A}^n_{\mathbb{R}}$ are algebraic subsets. Find specific examples of subsets that are not of the form $V(J)$. (Hint: we have already given one example of this in this chapter, if you get stuck, look through the examples and see which one is not algebraic and explain why.)*

1.3 Conclusions

We have just shown a condition for $V(I(X)) = X$ to be true. This condition is satisfied by algebraic sets. In the coming sections, when we have developed more general ring theory, we will be able to classify the special ideals of $\mathbb{R}[x_1, \ldots, x_n]$ so that the analogous property, $I(V(J)) = J$ will hold. This condition is a bit more subtle and to prove it, we will actually need to use the famous Nullstellensatz of Hilbert. Before we do that, however, we conclude this chapter by reviewing the relationships between $\mathbb{R}[x_1, \ldots, x_n]$ and $\mathbb{A}^n_{\mathbb{R}}$.

Consider the following map:

$$V : \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\} \longrightarrow \{\text{subsets of } \mathbb{A}^n_{\mathbb{R}}\}$$

We see that this map as written above is not surjective. Indeed, by Homework 9 not every subset of $\mathbb{A}^n_{\mathbb{R}}$ is of the form $V(J)$ for some ideal J. We also saw in this chapter that V is not injective as written either. As we computed, $V(x^2) = V(x)$ but $(x^2) \neq (x)$. So it would seem our maps are not very friendly at the moment. Now consider the map:

$$V : \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\} \longrightarrow \{\text{algebraic subsets of } \mathbb{A}^n_{\mathbb{R}}\}$$

Now although our map V is still not injective for the same reason as before, it is now surjective onto the new set of algebraic subsets of $\mathbb{A}^n_{\mathbb{R}}$. Injectivity will have to wait until the next chapter, however.

Remark: The critical reader might object to our apparent excitement over a surjective map. Indeed, we defined the set of algebraic subsets as the image of all ideals, and therefore, it's not incredibly interesting that by restricting the range of our function to the image of V that it is all of sudden surjective. The interested reader might find it amusing to reconstruct the maps V and I using not ideals of $\mathbb{R}[x_1, \ldots, x_n]$ but just arbitrary subsets. In doing so, he or she will discover (we actually proved it in this chapter!) that even under these conditions, $V(S)$ will still be an algebraic subset of $\mathbb{A}^n_{\mathbb{R}}$ and $I(X)$ will still be an ideal. Of course doing this gives no greater insight into what is really going on here, which is why this comment is headed with the word Remark instead of Homework.

Homework 10. *Now for an* actual *homework problem. We just discussed the properties of the function V with the range restricted to algebraic subsets. What can be said about the function*

$$I : \{\textit{algebraic subsets of } \mathbb{A}^n_{\mathbb{R}}\} \longrightarrow \{\textit{ideals of } \mathbb{R}[x_1, \ldots, x_n]\}?$$

Prove it is not surjective, but that it is injective. (Hint: In restricting everything down to algebraic subsets, we ominously have not used which particular recent fact?)

1.4 Basic Set Theoretic Exercises

These exercises are not necessarily related to algebraic geometry, but nevertheless they are important for everyone to know. If you are like me, then it is likely that you won't remember which theorem is which, but after you do these exercises, you'll see that it's not difficult to prove any of these statements so you can rederive the results you need as needed.

Homework 11. *Suppose that $f : X \to Y$ is injective and $g : Y \to Z$ is injective. Prove that $g \circ f : X \to Z$ is injective.*

Homework 12. *Same question for surjective maps*

Homework 13. *Suppose that $f : X \to Y$, $g : Y \to X$ and $f \circ g = Id_Y$, the identity map on Y. Then prove*
1) f is surjective;
2) g is injective.

Homework 14. *Formulate your own version of the previous problem if $g \circ f = Id_X$ instead. Prove all claims you make. (Hint, one "proof" might be to simply turn your paper upside down or tilt your head)*

Homework 15. *Reread Proposition 5 and the the section "Conclusion" of this chapter and see how these theorems on set theory immediately imply our conclusions and the homework problem.*

2 Commutative Rings

2.1 Basic Definitions

In the first chapter we discussed the polynomial ring $\mathbb{R}[x_1, \ldots, x_n]$. In the course of this book, we will mainly be dealing with polynomial rings, so this is the concept we'd like you you to keep in mind. It will be useful, however, to introduce the concept of an abstract commutative ring. If nothing else, it will be nice to show that the theorems we will be proving work over most rings and not just $\mathbb{R}[x_1, \ldots, x_n]$.

Definition 5. *A* ring *is a set R together with two binary operations, denoted $+ : R \times R \longrightarrow R$ and $\cdot : R \times R \longrightarrow R$, such that*

1. *$(a+b)+c = a+(b+c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$ (associative law)*
2. *$a + b = b + a$ for all $a, b \in R$ (commutative law)*
3. *There exists an element $0 \in R$ such that $a + 0 = a$ for all $a \in R$ (additive identity)*
4. *For all $a \in R$, there exists $b \in R$ such that $a + b = 0$ (additive inverse)*
5. *$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ for all $a, b, c \in R$ (distributive law)*
6. *There exists an element $1 \in R$ such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$. (multiplicative identity)*

If in addition, the multiplication $\cdot$ is commutative, $a \cdot b = b \cdot a$ for every $a, b \in R$ then we say that R is a commutative ring.

Examples: There are plenty of examples of rings that you are already familiar with.

- The integers, rational numbers, and real numbers $(\mathbb{Z}, \mathbb{Q}, \mathbb{R})$ are all commutative rings with identities 0 and 1.
- The ring of $M_n(\mathbb{R})$ of $n \times n$ matrices with real entries is a ring under matrix addition and multiplication. It is not commutative if $n > 1$. (Proof left to the reader)
- Let X be a set. Let $S_{\mathbb{R}}(X)$ denote the set of all real valued functions on X. This is a commutative ring. This is the set of all functions $f : X \to \mathbb{R}$. Addition and multiplication are defined in terms of the corresponding operations on the real numbers. For instance, if f, g are functions from X to $\mathbb{R}$ then $f + g$ is the function defined by $(f + g)(x) = f(x) + g(x)$, and $fg(x) = f(x)g(x)$. (We are simply adding and multiplying the real numbers $f(x)$ and $g(x)$.)

- The above example generalizes if $\mathbb{R}$ is replaced by any ring R. $S_R(X)$ is always a ring, and is commutative if R is commutative.
- The integers "modulo n" form a commutative ring denoted $\mathbb{Z}_n$ Addition and multiplication of integers is carried out as usual and then the result is reduce mod n.
- Any field is a commutative ring.

One thing that is different in the case of arbitrary rings versus, say, the integers is that we may multiply to get zero in non-trivial ways. For instance, in the ring $\mathbb{Z}_6$ we have that $3 \cdot 2 = 0$. Similar examples exist for matrices and many other rings. To discuss this phenomena, we give a few definitions.

Definition 6. *Let R be a commutative ring. We say that an element $x \in R$ is a zero-divisor if $x \neq 0$ and there exists some $y \neq= 0$ in R such that $xy = 0$.*

Definition 7. *Let R be a commutative ring. We say that an element $x \in R$ is nilpotent if there exists an integer n such that $x^n = 0$.*

Homework 16. *In the examples above, we can have the unfortunate problem of* nilpotent *elements. In the examples of $M_n(\mathbb{R})$, and $\mathbb{Z}_n$, find nonzero elements x and some positive integer m such that $x^m = 0$. If you have two nilpotent elements x, y is their sum nilpotent? Check both M_n and $\mathbb{Z}_n$ carefully...*

The previous homework problem should have familiarized the reader with these new concepts. A ring with no zero divisors is called an *integral domain* and this definition is equivalent to the following:

Definition 8. *A commutative ring R is called an integral domain if for some $x, y \in R$, $xy = 0$, then either $x = 0$ or $y = 0$. Equivalently, if $x, y \neq 0$ then $xy \neq 0$.*

In other words, the only way you can multiply to zero is by using zero. All of our favorite rings like $\mathbb{Z}, \mathbb{R}$, and all fields are integral domains. As a general rule, most of the basic rings we start with will be integral domains, but as we start passing to more sophisticated structures, adding zero-divisors actually makes things a bit easier in some sense. Don't worry so much now about all of this, it will come up again when we discuss quotient rings.

Homework 17. *In general we do not have a cancelation law for commutative rings. Indeed, $4 \cdot 9 = 4 \cdot 4$ in the ring of integers mod 10. But $9 \neq 4$. (We can't cancel the 4s). We do have such a law for integral domains. Prove that if R is an integral domain and $xz = yz$ for some $z \neq 0$, then $x = y$.*

Homework 18. *Prove that $\mathbb{R}[x]$ is an integral domain.*

In doing the previous homework assignment, what did you notice about the proof? It most certainly didn't rely on any particular facts about $\mathbb{R}$ other than the fact that $\mathbb{R}$ is an integral domain. Indeed, it seems time that we stepped out a bit further and defined a polynomial ring in general.

Definition 9. *Let R be a commutative ring. Then the polynomial ring $R[x_1, \ldots, x_n]$ is the set of all polynomials in n variables with coefficients in R. Addition and multiplication is defined in the usual way.*

Remark: When we say "the usual way" in the above definition, don't feel as if we are only telling you half of the story. Indeed, writing out the explicit formula for multiplication, (or even for the general element of $R[x_1, \ldots, x_n]$ for that matter!) would not lead to any greater understanding of the polynomial ring. Just think of it as $\mathbb{R}[x_1, \ldots, x_n]$ with generalized coefficients and you will be fine.

Note that since by definition, each element of R is a constant polynomial, it is an element of $R[x_1, \ldots, x_n]$. Thus we can consider $R \subset R[x_1, \ldots, x_n]$. With this in mind, if R is not an integral domain, then neither is $R[x_1, \ldots, x_n]$ since zero divisors in R are naturally zero divisors in $R[x_1, \ldots, x_n]$. This along with a slight tweaking of Homework 18 we have the following:

Proposition 6. *$R[x_1 \ldots, x_n]$ is an integral domain if and only if R is an integral domain.*

Proof. Note that, for example, $R[x, y] = (R[x])[y]$. Just think of $R[x, y]$ as polynomials in y with coefficients that are polynomials in x. (If this makes you wary, please assign yourself an extra homework assignment and convince yourself that this is true.) Thus from our tweaked homework problem, we know that if R is an integral domain, then so is $R[x]$. Thus by the same homework problem, since $R[x]$ is an integral domain, so is $(R[x])[y] = R[x, y]$. Thus continuing by induction, our proof is complete. □

For the remainder of these notes, we assume all rings are commutative unless stated otherwise. For most applications that we do, the rings will mostly be polynomial rings over fields, or quotients thereof. To define a quotient ring, however we need to develop an abstract version of the polynomial ideal introduced in the first chapter.

Definition 10. *Let R be a ring. A subset $I \subset R$ is called an ideal if I is nonempty and the following two properties hold:*

- *If $x, y \in I$ then $x + y \in I$. (Closure under addition)*
- *If $x \in I$ and $r \in R$ (any element in R) then $rx \in I$. (closure under ring multiplication)*

Just as we saw in the first chapter, there are many different examples of ideals in the polynomial ring $\mathbb{R}[x]$. Of course all these ideals have their counterparts in arbitrary rings $R[x]$ as well. Now let's look at some ideals in other rings

Example:

- Let R be a ring, and let $x \in R$. By (x) we mean the ideal of all multiples of x:
$$(x) = \{rx \ : \ r \in R\}.$$

This is clearly closed under ring multiplication. To see it is closed under addition, consider $rx + sx = (r + s)x \in (x)$.

- Let R be a ring, and $x_1, \ldots, x_n \in R$. Then we define

$$(x_1, \ldots, x_n) = \{r_1 x_1 + \ldots r_n x_n \ : r_i \in R\}$$

just as we did for polynomial rings.

- Let $R = \mathbb{Z}_6$, a ring with 6 elements $\{0, 1, 2, 3, 4, 5\}$ and addition is done modulo 6. Consider $I = (2)$. Then it is evident that

$$(2) = \{0, 2, 4\}$$

- Let $R = \mathbb{Z}_6$ again, only this time consider (5). Check for yourself that $(5) = R$.

What happened in the last example, where an element generated the entire ideal, is an important notion.

Definition 11. *Let R be a ring. Then $x \in R$ is called a unit if there exists $y \in R$ such that $xy = 1$.*

In the last example, 5 is a unit because $5 \cdot 5 = 1$. In fact we have the following:

Homework 19. *Prove that in a ring R, x is a unit if and only if $(x) = R$.*

Definition 12. *A field is a commutative ring such that every nonzero element is a unit.*

Note that this definition and the preceding homework problem give a very nice result about ideals of a field.

Proposition 7. *Let F be a field. Then the only ideals of F are $\{0\}$ and F.*

Proof. Let I be a nonzero ideal. Then there exists $x \in I$ with $x \neq 0$. By the above homework, $(x) = F$. But then $F = (x) \subset I$ so $I = F$. □

Homework 20. *Let F be a field, and $F[x]$ the ring of polynomials over F. Prove that if $f \in F[x]$ is a unit then $f \in F$. Thus, no nonconstant polynomial can be a unit.*

To see the necessity that F be a field in the previous problem, consider the following example. Let $R = \mathbb{Z}_4$ and consider $R[x]$. Then in $R[x]$,

$$(2x + 1)(2x + 1) = 4x^2 + 4x + 1 = 0 + 0 + 1 = 1$$

so that $2x+1$ is a unit. For a very challenging homework problem, try to classify all units in $R[x]$. To see how to do this for a general ring R, see *Introduction to Commutative Algebra* by Atiyah & MacDonald.

Homework 21. *Compute the units in $\mathbb{Z}_6, \mathbb{Z}_8, \mathbb{Z}_{12}, \mathbb{Z}_n$. (Hint for $\mathbb{Z}_n$: Let a, b be integers. Then what numbers can be expressed as $ax + by$ where y, x are integers? A classic result from number theory says that you can write $gcd(a, b)$ in this form as well as all of its multiples, and that these are the only solutions.)*

Homework 22. *Let F be a field. Prove F has no zero-divisors.*

2.2 Quotient Rings: Preview

So far, the contents of this chapter have been pretty elementary. A ring is simply something that behaves like the integers, or like the set of polynomials. Ideals are simply specialized subsets of rings, and units are just elements with multiplicative inverses. The next topic we will discuss is the notion of a quotient ring. We start this section with a motivation from modular arithmetic.

We have alluded to modular arithmetic many times already. We assumed the reader was at least familiar with the operations thereof. (Add and subtract as usual, and then take the remainder when you divide by n.) For example, in $\mathbb{Z}_{12}$, $6 \cdot 7 = 42 = 3 \cdot 12 + 6$ so $6 \cdot 7 = 6$ in $\mathbb{Z}_{12}$. What the reader may not be familiar with, is the more formal construction of modular arithmetic that we now give.

Modular Arithmetic:

Definition 13. *Let S be a set. An equivalence relation $\sim$ is a relation on S such that the following hold*

- *$x \sim x$ for all x (Reflexive Property)*
- *If $x \sim y$ then $y \sim x$ for all x, y (Symmetric Property)*
- *If $x \sim y$, $y \sim z$ then $x \sim z$ for all x, y, z (Transitive Property)*

(Strictly speaking, we should define what we mean by a relation, but doing so is clunky. Instead just think of $\sim$ as a new way of saying elements are equal.

You already know many examples of equivalence relations from previous math courses. For example, if S is the set of triangles in the plane, then if we define

$$T \sim T' \text{ if } T \text{ and } T' \text{ are congruent triangles}$$

then $\sim$ is an equivalence relation.

For another example, let S be the set of all people on Earth. Then if we define

$$X \sim Y \text{ if } X \text{ and } Y \text{ are related by blood}$$

then $\sim$ is an equivalence relation.

Homework 23. *The properties in the definition, (reflexive, symmetric, transitive) do not imply each other. Play around a bit and try to find relations (these won't be equivalence relations by definition) that have exactly one of the properties, or exactly two of the properties.*

Now let $S = \mathbb{Z}$ be the set of integers. Let n be a positive integer. Then we define $\sim$ as follows:

$$x \sim y \text{ iff } x - y \text{ is a multiple of } n \ .$$

Using our notation of ideals, we could rewrite this as

$$x \sim y \text{ iff } x - y \in (n).$$

To see that this is an equivalence relation is easy, and we will only show the transitive property. Suppose that $x \sim y$ and $y \sim z$. Then $x - y = kn,\ y - z = ln$ for some $k, l \in \mathbb{Z}$. Then $x - z = kn + ln = (k + l)n \in (n)$, so $x \sim z$.

Now that we have an equivalence relation we can define the *equivalence class* of an element.

Definition 14. *Let S be a set, and $\sim$ an equivalence relation on S. Then if $x \in S$ we define*

$$\overline{x} = \{y \in S \ : \ x \sim y\}.$$

To give this some context, in the example of triangles, $\overline{T}$ is the set all of triangles congruent to T. In the set of all people, $\overline{\text{Kevin Bacon}}$ is the set of all people related to Kevin Bacon. (If you believe in creationism, then for any people X and Y, $\overline{X} = \overline{Y} = \{\text{the whole world}\}$ since everyone is related.)

In our final example, if $n = 5$ then $\overline{3} = \{\ldots, -7, -2, 3, 8, 13, \ldots\}$ or the set of all numbers 3 mod 5. Thus $\overline{3} = \overline{8} = \overline{-2}$.

Definition 15. *Let $S = \mathbb{Z}$ be the set of integers and let $\sim$ be the equivalence relation defined above. We define the "integers modulo n" to be the set of equivalence classes $\overline{x}$ with the operations $\overline{x} + \overline{y} = \overline{x + y}$, and $\overline{x}\,\overline{y} = \overline{xy}$.*

The reader should now check that addition and multiplication of these equivalence classes is well-defined. (That is, that whether you pick $\overline{3}$ or $\overline{8}$ to represent the class of 3 mod 5, the answer will not depend on this.) This definition of modular arithmetic turns out to be the same as the one we are used to, but was constructed in a more general way. When we write

$$21 + 25 = 2 \text{ mod } 11$$

we are really adding $\overline{21}$ and $\overline{25}$ and getting $\overline{1}$, but we drop the bar notation because there really is no point of confusion.

We briefly review what we just did because it is important: We took our set to be the ring of integers. We took a subset, an ideal (n) and then said $x \sim y$ if $x - y \in (n)$. Then we took the class of all equivalence classes, $\overline{x}$. In this new set we had the cool property that $\overline{n} = \overline{0}$ so that basically n now becomes 0. Keep these things in mind as we move on to the more general case.

2.3 Quotient Rings

Let I be an ideal of a ring R. Define an equivalence relation $\sim$ on R as follows:

$$x \sim y \text{ iff } x - y \in I.$$

Let $\overline{x}$ denote the equivalence class of an element x. We ask now, what do elements equivalent to x look like?

Suppose that $x \sim y$. Then $x - y \in I$ so that $x - y = a$ for some $a \in I$. Thus $y = x - a$, which is x plus some element in I. In fact it is not hard to generalize this to the following

Claim: $\overline{x}$ is the set of all elements of the form $x + a$ where $a \in I$.

For this reason, rather than writing $\overline{x}$ for the equivalence class of x, we simply write $x + I$. We can now talk about the set of all equivalence classes, and we call this set R/I. Namely,

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

We now try our luck at defining addition and multiplication on R/I and hope that we get a ring. There's really only one way to do this, so we try it:

$$(a + I) + (b + I) = (a + b + I),$$

$$(a + I) \cdot (b + I) = (ab + I).$$

Looking at this definition it seems to have all the properties of a ring that we want. The identity elements are $0 + I$ and $1 + I$, and all the distributive properties simply come from the corresponding facts about R. The only problem is of course the fact that these operations might not be well defined. We will prove this immediately after stating our result.

Proposition 8. *With the above operations defined, the set R/I becomes a commutative ring, which we call the quotient ring of R by I. (Sometimes read R mod I.)*

Proof. We must check that addition is well defined. Suppose that we are adding $a+I$ and $b+I$, but that we know $a+I = c+I$ and $b+I = d+I$. Then we must show that our formula for addition doesn't depend on which representation we choose. Thus we must show that

$$(a + I) + (b + I) = (c + I) + (d + I)$$

Which is the same as showing $a + b + I = c + d + I$. The first of these is the equivalence class of $a + b$ and the second of $c + d$. These two sets will be equal precisely if $a + b$ and $c + d$ are equivalent, that is, if

$$a + b - (c + d) \in I$$

But $a + b - (c + d) = (a - c) + (b - d)$ and since $a \sim c$ and $b \sim d$ we have $a - c$, $b - d \in I$, and the result follows.

For multiplication: Suppose that a, b, c, d are as above. Then we know that $a \sim c, b \sim d$, so $a - c \in I$, $b - d \in I$. Since I is an ideal

$$b(a-c) + c(b-d) = ab - cd \in I.$$

So that $ab \sim cd$ and thus $ab + I = cd + I$. □

Remark: It will greatly benefit the reader to see the following: In the quotient ring R/I,

$$x + I = y + I \text{ iff } x - y \in I.$$

This is just a restatement of the definition we gave above, but we feel this gives a better feel for what is going on.

Examples:

1. $\mathbb{Z}_n$ is the quotient ring $\mathbb{Z}/(n)$. In this example, what we formerly called $\overline{x}$ is now $x + (n)$.

2. Let R be any ring, and let $I = R$. Then $R/I = 0$ since $x + I = y + I$ for any two elements in R. (Really, it's just saying that $x - y \in I = R$, which is a pretty easy definition to satisfy.)

3. Let $R = \mathbb{R}[x]$, $I = (x^2)$. We make the claim that

   $$R/I = \{a + bx + I \ : a, b \in R\}.$$

   Indeed, suppose that $f(x) = a_n x^n + \cdots + a_1 x + a_0$ is any element of R. Then

   $$f(x) - (a_1 x + a_0) = x^2(a_n x^{n-2} + \cdots + a_2) \in (x^2)$$

   so $f(x) + I = a_0 + a_1 x + I$. In other words, once we took the quotient by x^2, our resulting ring was still a polynomial ring, only this time $x^2 = 0$.

The idea hit on in the last example is precisely the way to think about quotient rings. *If you quotient by an ideal I, then the resulting ring is the same, except now every element of I is* 0. Indeed, in $\mathbb{Z}_n = \mathbb{Z}/(n)$, all multiples of n are considered 0. In R/R, everything is considered 0.

Example: We now do a slightly harder example, and instead of being completely rigorous, we appeal to what we just discovered. Let R be any ring, and let $I = (x^2 - 5)$ be an ideal of $R[x]$. Then in $R[x]/I$, we have the relation $x^2 - 5 = 0$, or in other words, $x^2 = 5$. Thus whenever we want to write an equivalence class of a polynomial, we first write it down and then note that $x^2 = 5$. It isn't hard to see then, that everything reduces down nicely to just linear functions. For example,

$$x^4 + 3x^3 + x + I = (x^2)^2 + 3(x^2)x + x + I = 5^2 + 3(5)x + x + I = 25 + 16x + I.$$

We might even get lazy and leave the "$+I$" out and just say that

$$x^4 + 3x^3 + x = 25 + 16x \bmod (x^2 - 5)$$

just as we did with modular arithmetic. To complete this example, we really should explain how to multiply. This can be done as follows: In $R[x]/I$,

$$(ax+b)(cx+d) = (5ac+bd) + (ad+bc)x \bmod (x^2-5).$$

Basically the rule is: multiply as normal, and then use the relation $x^2 = 5$ to simplify.

Homework 24. *Consider the following subset of the real numbers,*

$$\mathbb{Z}[\sqrt{5}] = \{a + b\sqrt{5} \ : \ a, b \in \mathbb{Z}\}.$$

Prove that S is a commutative ring and write out the multiplication rule.

Now consider the last example, with $R = \mathbb{Z}$. In the end we computed a ring $\mathbb{Z}[x]/I$ where we constructed a "square root" of 5. In other words, we made x satisfy $x^2 = 5$. Convince yourself that the formal construction in the example is the same as this more down to earth problem.

It is often better to use the more abstract version, because we cannot always write down an explicit formula for the solutions to some equations. For example, the solution in radicals to $x^3 - 3x^2 + x - 5 = 0$ is probably very nasty, and a result from Galois Theory tells us that if the degree is 5 or greater then in general, the solution cannot *be expressed in terms of radicals.*

Homework 25. *Let k be a field. What is $k[x]/(x)$.?*

Homework 26. *Let k be a field, and $R = k[x, y]$. What is $R/(x, y)$? $R/(x^2, y^3)$?*

Homework 27. *What is $\mathbb{R}[x]/(x^2+1)$? Write out the rule for multiplying elements, does this object look familiar to you?*

3 Applications to Algebraic Geometry

In the first section of these notes we introduced two maps

$$V : \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\} \longrightarrow \{\text{subsets of } \mathbb{A}^n_{\mathbb{R}}\}$$

$$I : \{\text{subsets of } \mathbb{A}^n_{\mathbb{R}}\} \longrightarrow \{\text{ideals of } \mathbb{R}[x_1, \ldots, x_n]\}$$

and we saw that these maps are not quite bijective. If restrict the subsets of $\mathbb{A}^n_{\mathbb{R}}$ to be algebraic, then we are halfway there. We are now better to discuss the other side of the problem; the one of ideals.

Definition 16. *Let R be a ring, and I an ideal. The radical of I denoted $\sqrt{I}$ is given by*

$$\sqrt{I} = \{x \in R \ : \ x^n \in I \text{ for some } n\}.$$

An ideal I is called a radical ideal if $I = \sqrt{I}$.

Example: Let $R = \mathbb{R}[x_1, \ldots, x_n]$, and let $I = (x^3)$. Then $\sqrt{I} = (x)$.

Proof. Clearly $(x) \subset \sqrt{(x^3)}$, since if $x \cdot f(x) \in (x)$, then $(x \cdot f(x))^3 \in (x^3)$. Conversely, suppose that $f(x) \in \sqrt{(x^3)}$. Then $f(x)^n \in (x^3)$ for some n. Thus $f(x)^n$ has a factor of x^3, so f has a factor of x, thus $f(x) \in (x)$. □

Homework 28. *Concoct your own proof that* $\sqrt{(x^n)} = (x)$.

Homework 29. *In an integral domain prove that* $\sqrt{\{0\}} = \{0\}$. *Show this is not true for general rings. Finally, show that if* $\sqrt{I} = R$, *then* $I = R$.

The radical in the previous homework has a special name and merits its own definition.

Definition 17. *The nilradical of a ring* R, $Nil(R)$ *is the radical of the zero ideal,* $\sqrt{0}$.

Proposition 9. *The radical of an ideal is an ideal.*

Proof. Let I be an ideal in R, and $x, y \in \sqrt{I}$. (Say $x^n \in I, y^m \in I$.) If $r \in R$ then clearly $(rx)^n = r^n x^n \in I$ so $rx \in \sqrt{I}$. Finally we must show that $x + y \in \sqrt{I}$. This can be done by taking high enough powers. By the binomial theorem,

$$(x+y)^{m+n} = \sum_{i=0}^{m+n} r_i x^i y^{m+n-i}$$

for some coefficients $r_i \in R$. Now look at each of these terms. If $i \geq n$, then $x^i \in I$, and thus so is the ith term. If $i < n$, then $m+n-i > m$ so $y^{m+n-i} \in I$. Thus every term is in I, so $x + y \in \sqrt{I}$. □

To give us more practice with quotient rings we, prove the following proposition relating the radical of an arbitrary ideal with the nilradical of a quotient ring.

Proposition 10. *Let* I *be an ideal of a ring* R. *Then* $\sqrt{I}$ *can be identified with* $Nil(R/I)$ *in the natural way:* $x \mapsto x + I$.

Proof. Let $x \in \sqrt{I}$. Thus $x^n \in I$ for some n. We want to show that $x + I \in \mathrm{Nil}(R/I)$. But $(x+I)^n = x^n + I = 0 + I$ so $x + I \in \mathrm{Nil}(R/I)$.

Conversely, if $y + I \in \mathrm{Nil}(R/I)$, then $y^n + I = 0$ for some n, so that $y^n \in I$ and $y \in \sqrt{I}$ as required. □

3.1 Hilbert's Nullstellensatz

In this section we will state and prove the majority of Hilbert's Nullstellensatz. In German, the name literally means "Theorem on the zeros of a polynomial". Using the Nullstellensatz, we will be able to better answer the questions posed in the first section about the relationship between the maps I and V. Throughout this section we will do all of our work over an abstract field k. The affine plane $\mathbb{A}^n_k$ should be identified with k^n if it makes you feel more at ease. We begin with a small result:

Proposition 11. *Let X be a subset of $\mathbb{A}^n_k$. Then $I(X)$ is a radical ideal.*

Proof. It is always true that $I \subset \sqrt{I}$ for any ideal I (since $f^1 \in I$ if $f \in I$), so we must show the other inclusion. Suppose $f \in \sqrt{I(X)}$. Then for some integer n, $f^n \in I(X)$. This means that $f^n(x) = 0$ for each $x \in X$. But this means that $f(x) = 0$ for each $x \in X$, so $f \in I(X)$. □

In the last proposition we saw that $I \subset \sqrt{I}$. In general we can have ideals that contain one another. For example, in $k[x]$, we have

$$(x) \supset (x^2) \subset (x^3) \subset \cdots (x^n) \supset \cdots$$

In $\mathbb{R}[x]$ we have $(x^2-1) \supset (x-1)$ and $(x^2-1) \supset (x+1)$. This is true since any multiple of x^2-1 is also a multiple of $x-1$ and $x+1$. Thus we begin to see that the inclusion of ideals is somehow related to factorization. So far, all the ideals we have listed in this section are ones generated by a single element. These ideals are given a special name:

Definition 18. *A principal ideal is an ideal generated by a single element, that is, one of the form (f). If every ideal in a ring R is principal, then we say R is a principal ideal domain. (PID)*

Note that just because an ideal is presented to us with more than one generator this does not mean that the ideal is not principal. For example, in $\mathbb{R}[x]$ consider $(x, 2x)$. This ideal is clearly the same as (x). We say the extra generator is *redundant*. To give this some comparison to linear algebra, recall that a set of n vectors doesn't have to span an $n-$dimensional space. The span of the vectors could easily be the same as the span of just one of them.

Finally, we note that not all rings are PIDs. Indeed, consider the polynomial ring $\mathbb{R}[x, y]$. Consider the ideal (x, y). This cannot be generated by one element.

Homework 30. *Prove that (x, y) is not a principal ideal.*

PIDs are extremely important in algebra. In addition to the obvious ease that all ideals being principal gives, these rings also have many other properties that we will not discuss here. Thus it is important to have a nice supply of these objects.

Proposition 12. *Let k be field. Then $k[x]$ is a PID.*

Proof. We sketch a proof of this. Let I be an ideal in $k[x]$. Then I has a set of generators, $(f_1, \ldots, f_n)$. Let g be the g.c.d. of the the polynomials. Then $I = (g)$. It is a standard fact from the Euclidean algorithm that g can be written as a combination of the f_i, so $(g) \in I$. And also, any element of I is a sum of multiples of the f_i so that it is also a sum of multiples of g, the g.c.d. □

We now go back to our example with (x^2-1). We saw that for example, $(x^2-1) \subset (x-1)$.

Question: In $\mathbb{R}[x]$ can you find an ideal that properly contains $(x-1)$ that is not equal to the whole ring?

The answer to this question is no. Suppose that there were some ideal J with this property, and $(x-1) \subset J$ strictly. Then there is some $f \in J \backslash (x-1)$. This f is not a multiple of $(x-1)$ so we use polynomial long division to get

$$f(x) = (x-1)q(x) + r$$

where r is a constant and not equal to 0. But then

$$1 = \frac{1}{r}f(x) - \frac{1}{r}(x-1)q(x) \in J$$

so $1 \in J$ and thus $J = \mathbb{R}[x]$.

In general, if the only ideal properly containing I is the whole ring, then I is called a maximal ideal.

Definition 19. *An ideal $m \subset R$ is maximal if*

$$m \subset J \subset R$$

implies either $J = m$ or $J = R$.

(It is standard to use the letter m to denote a maximal ideal)

The work we did above easily generalizes to the following proposition:

Proposition 13. *Let k be a field. In $k[x]$, ideals of the form $(x-a)$ are maximal.* □

In this discussion, there is an inherent problem in the field that we are working over. For instance, in $\mathbb{R}[x]$, the polynomial x^2+1 does not factor, and is therefore a maximal ideal. (The details are left to the reader). But in $\mathbb{C}[x]$, this polynomial factors as $(x+i)(x-i)$ so the ideal is it generates is not maximal. These difficulties are annoying so we'll just eliminate them for the rest of the section as follows.

Definition 20. *A field k is called algebraically closed if every polynomial has a root. (This definition is equivalent to saying that every polynomial factors into a product of linear terms)*

Proposition 14. *If k is algebraically closed, then the only maximal ideals of $k[x]$ are of the form $(x-a)$.*

Proof. We have already seen that all ideals of this form are maximal. So now suppose that we have an ideal m that is maximal, but not of this form. Then since $k[x]$ is a PID, $m = (f)$ for some $f \in k[x]$. Then if f is not a linear polynomial, it has a linear factor, $(ax-b)$ and then $m = (f) \subset (ax-b)$ where the inclusion is strict, so m is not maximal. □

The fact that $k[x]$ is a PID was very crucial in our proof of the above proposition. Equally important was the fact that k is algebraically closed. For example, in $k[x,y]$, the ideal $(x-a)$ is no longer maximal because it is contained in $(x-a, y)$, which is not a principal ideal. The maximal ideals, however behave very nicely and the work we did above generalizes nicely to the following proposition

Proposition 15. *(Nullstellensatz Part 1): Let k be an algebraically closed field. Then each maximal ideal of $k[x_1, \ldots, x_n]$ has the form*

$$(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n).$$

Proof. This is the first part of Hilbert's famous Nullstellensatz. We do not give a full proof here because the insight is the same as in the case in one variable and the proof is technical. For a reference see Reid, or Artin. □

Corollary 1. *There is a natural bijection between maximal ideals of $k[x_1, \ldots, x_n]$ and points $(a_1, \ldots, a_n)$ in $\mathbb{A}^n_k$.*

Proposition 16. *(Nullstellensatz Part 2): Let I be a proper ideal of $k[x_1, \ldots, x_n]$. Then $V(I) \neq \emptyset$. (In other words, if $I = (f_1, \ldots, f_r)$ then the system of polynomial equations $f_i = 0$ has a solution.)*

Proof. We will use the following fact: Every ideal is contained in a maximal ideal. Thus $I \subset m = (x_1 - a_1, \ldots, x_n - a_n)$ for some $(a_1, \ldots, a_n)$. Each element f of I is a combination of $x_i - a_i$,

$$f = \sum c_i(x_i - a_i).$$

It is then clear that $f(a_1, \ldots, a_n) = 0$ so I vanishes at $(a_1, \ldots, a_n)$. □

Proposition 17. *(Nullstellensatz Part 3): $I(V(J)) = \sqrt{J}$.*

Proof. We first show that $\sqrt{J} \subset I(V(J))$. Let $f \in \sqrt{J}$. Then for some n, $f^n \in J \subset I(V(J))$. But $I(V(J))$ is a radical ideal since it is of the form $I(X)$. Thus $f \in I(V(J))$.

The heart of this proof lies in showing the other inclusion. We assume that $f \in I(V(J))$. We must show that $f^m \in J$ for some integer m. Let $J = (f_1, \ldots, f_r)$. Then consider

$$\widetilde{J} = (f_1, \ldots, f_r, 1 - yf) \subset k[x_1, \ldots, x_n, y].$$

Claim: $V(\widetilde{J}) = \emptyset$.

To show this we pick an arbitrary point $a = (a_1, \ldots, a_{n+1}) \in \mathbb{A}^{n+1}_k$. We will show that $\widetilde{J}$ does not vanish at a.

Case 1: $(a_1, \ldots, a_n)$ is a common zero of $f_1, \ldots, f_r$. Then $(a_1, \ldots, a_n) \in V(J)$ and since $f \in I(V(J))$, $f(a_1, \ldots, a_n) = 0$ as well. Thus

$$(1 - yf)(a) = 1 - y \cdot f(a_1, \ldots, a_{n+1}) = 1 - 0 = 1$$

so $\widetilde{J}$ does not vanish at a.

Case 2: If $(a_1, \ldots, a_n) \notin V(J)$ then some $f_i(a_1, \ldots, a_n) \neq 0$. But then $f_i(a_1, \ldots, a_{n+1}) \neq 0$ as well since f_i does not depend on y.

Thus $V(\widetilde{J}) = \emptyset$ which by the Nullstellensatz Part 2 implies that

$$\widetilde{J} = (1) = k[x_1, \ldots, x_n, y].$$

Hence we can write 1 as a combination of the generators of $\widetilde{J}$.

$$1 = \sum_{i=1}^{r} g_i(x_1, \ldots, x_n, y) f_i + h(x_1, \ldots, x_n, y)(yf - 1).$$

This is an identity, so we are free to let y be anything we want. We set $y = 1/f$, to get

$$1 = \sum_{i=1}^{r} g_i(x_1, \ldots, x_n, 1/f) f_i + h(x_1, \ldots, x_n, y)(0).$$

Multiplying through by a high enough power of f to clear all the fractions, we get

$$f^n = \sum_{i=1}^{r} G_i(x_1, \ldots, x_n) f_i \in J$$

where the G_i are the just the resulting polynomials we get after the multiplication. □

We have thus just answered one of the first main questions of these notes: when do the operations V and I undo each other? We saw in the first chapter that $V(I(X)) = X$ if X is an algebraic set. Just now we saw that $I(V(J)) = \sqrt{J}$ so that if J is a radical ideal, then $I(V(J)) = J$. We remind you that this matches what we discovered experimentally before, that $I(V(x^2)) = (x) = \sqrt{(x^2)}$. Thus we finally have our bijection:

$$\{\text{radical ideals of } \mathbb{R}[x_1, \ldots, x_n]\} \longleftrightarrow \{\text{algebraic subsets of } \mathbb{A}^n_{\mathbb{R}}\}.$$

4 Bézout's Theorem and Projective Space

In this section we will begin by stating Bézout's Theorem for the affine plane and then use this to motivate Projective space and the general form of Bézout's Theorem.

For the first part of this section we will work over $\mathbb{R}^2$, mainly so that the graphs we illustrate for you are familiar from high school analytic geometry. It is actually easier (as it usually is) to think of everything as being over $\mathbb{C}^2$, since $\mathbb{C}$ is algebraically closed. Unfortunately, when we think of $\mathbb{C}$ we usually think of it as a plane, so that $\mathbb{C}^2$ often becomes a 4-dimensional space in our minds instead of a 2-dimensional one. If this chapter should teach you anything, it is that just because you cannot visualize something over an arbitrary field, or in the new projective space, you should not be afraid to get your hands dirty and discover the properties of the object.

4.1 Bézout's Theorem for $\mathbb{R}^2$

In high school we looked at certain curves in the plane. Usually these curves were complicated polynomial expressions and we determined such properties as

concavity, relative and global extrema, etc. Within this large class of curves, we also studied the *Conic Sections*. This class of curves includes the parabola, hyperbola and ellipse. The general form for the equation of a conic is

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0.$$

To simplify things a bit, we can make a change of coordinates using a method known as "rotation of axes" to eliminate the xy term. If $A, B \neq 0$, then we can complete the squares in the variables x and y, to get something of the form.

$$a(x-h)^2 \pm b(y-k)^2 = 1.$$

If the sign is a + then we get an ellipse (or circle). If the sign is negative, then we get a hyperbola. If A or C is zero, then we have a parabola.

Remark: In each of the cases listed above, very bad things could occur. For instance, the equation $x^2 + y^2 = 0$ represents only one point. The "hyperbola" $(x-y)(x+y) = 0$ represents two intersecting lines. Be aware that things like this exist and that our write-up above was not entirely accurate. Then ask yourself if you really wanted to read a write-up that was. If yes, do the following homework problem. If no, skip it.

Homework 31. *Write a proper write-up of the conic sections considering all special cases.*

Bézout's Theorem is a statement concerning the number of points of intersection of curves. Before we can state it, we will look at some examples using our conics. Consider a line L and a conic C. In how many points can L and C intersect? Figure 1 shows that this number can be 0, 1 or 2.

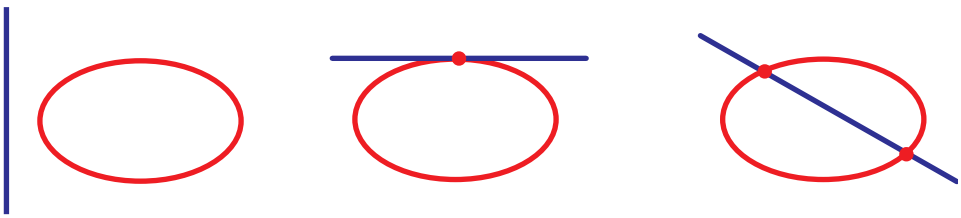


Figure 1: A conic intersecting a line

Consider two conics C and D. In how many points can these intersect. Figure 2 shows us that this number can be $0, 1, 2, 3, 4$. The number of possible intersections has increased. Why? The reason for this is because conics are degree two curves, while lines are degree one. The more degree, the more intersections! We make all this formal in the following definition and theorem.

Definition 21. *Let C be a curve given by a polynomial equation of the form $G(x, y) = 0$. Then the degree of C is the degree of the highest monomial term of G. (For example, the equation $x^5y + 2x + 9 = 0$ has degree 6.)*

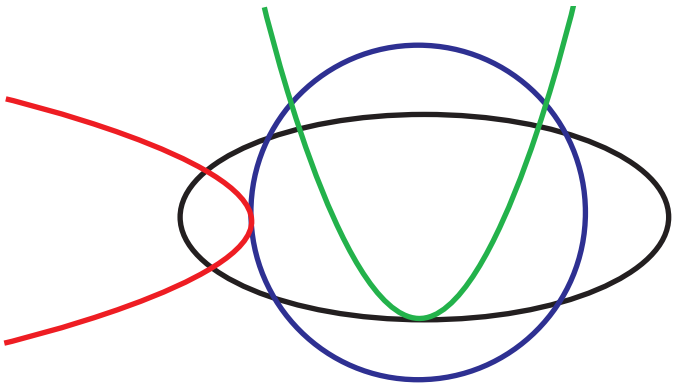


Figure 2: Two conics intersecting in 0,1,2,3 and 4 points

Proposition 18. *(Bézout's Theorem for $\mathbb{R}^2$) Let C and D be two curves in $\mathbb{R}^2$ of degrees c and d respectively. If C and D have no common component, then C and D intersect in at most cd points.*

We may prove Bézout's Theorem at a later time, and if we do, we will prove the general statement for the projective plane. We now try to motivate this statement and its generalization.

First things first, we must must abandon our notion of what polynomial curves look like. In high school we all encountered many different graphs of cubic polynomials, but what's important to remember was that all of these graphs were of the form

$$y = ax^3 + bx^2 + cx + d.$$

No terms involving higher powers of y were used.

As a first example, consider the equation

$$(y-1)(y-2)(x-3) = 0.$$

This is a cubic curve since the highest degree term is xy^2. But it is easy to see that this curve is the intersection of three lines. This is an example of what is called a *reducible cubic*.

Definition 22. *A polynomial curve C is said to be reducible if it can be written as a union $C = D \cup E$ where D and E are both polynomial curves.*

This definition of reducible is very geometric. In terms of algebra, it is good to associate reducible curves with polynomial expressions which factor.

Let us now generate a different type of reducible cubic curve. Take your favorite conic. Mine is the unit circle given by

$$x^2 + y^2 - 1 = 0.$$

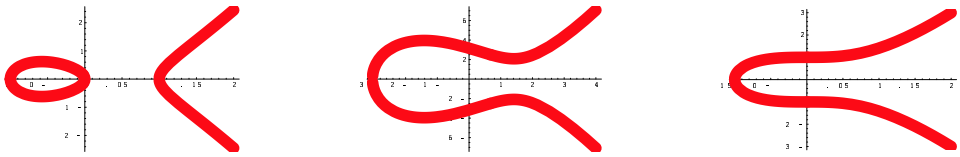
This has degree two, so let's multiply it by something of degree 1, a line. We'll pretend that I have a favorite line and that it is $x + y - 1 = 0$. Then when we take the product we get

$$(x + y - 1)(x^2 + y^2 - 1) = 0$$

which is a cubic polynomial equation. This curve is simply the union of our favorite circle and line.

Homework 32. *Let C be a reducible cubic, prove that C is either three lines, or an irreducible conic union a line. (Recall that the union of two lines is a reducible conic)*

Irreducible cubics come in all varieties. (No pun intended!) We illustrate a few in Figure 3.



$y^2 = x(x-1)(x+1)$ $\quad$ $y^2 = x^3 - 6x + 10$ $\quad$ $y^2 = x^3 + 1$

Figure 3: Some irreducible cubic curves

Now it should be clear what we mean when we say that two curves share a component. For example, the two conics

$$(x^2 - y)(x - 1) = 0 \quad \text{and} \quad (x^2 + y)(x - 1) = 0$$

both share the line $x - 1 = 0$. Therefore they intersect in infinitely many points, so Bézout's Theorem cannot apply.

4.2 A Closer Look at Bézout's Theorem

Perhaps the words "at most" stuck out to you when you read the statement of Bézout's Theorem . It would be great if we could say in precisely how many points two curves intersect. Of course we know that is a problem. Two lines can intersect or be parallel; a line and a circle can intersect anywhere from 0 to 2 times; the case for conics is seemingly even more complicated. We study the case of a circle and a line.

Consider the circle given by $x^2 + y^2 - 9 = 0$ and the line given by $x - 5 = 0$. Then Bézout's Theorem says that these two curves should intersect in at most 2 points. But these do not intersect at all in $\mathbb{R}^2$, since the curves are disjoint. Over the complex numbers, however, they intersect at the two points $(5, 4i)$ and $(5, -4i)$. So over the complex numbers the upper bound from Bézout's Theorem theorem actually gives the correct number of intersection points. This

suggests that to properly count everything, we should work over an algebraically closed field.

We have a different problem, however with the circle $x^2 + y^2 - 1 = 0$ and line $x - 1 = 0$. These intersect only at the point $(1, 0)$ no matter what field we are working over. Note however that in arriving at this solution we would have the equation $y^2 = 0$, so we see that $y = 0$ is a "double root". Thus if consider this *multiplicity*, we have two roots.

Definition 23. *Let $f(x)$ be a polynomial, and let $f(a) = 0$. Then the multiplicity of the root a is defined to be the highest power of $x - a$ that divides f.*

To recap, we learned that to properly count the number of points of intersection we should work over an algebraically closed field and consider multiplicity. Unfortunately, this is not enough as the case of parallel lines illustrates. Where do the lines $x = 1$ and $x = 2$ intersect? If you are tempted to say "at infinity" then you are exactly right. In what follows we carefully define a mathematical framework so that we can talk about "infinity" as if it were any other point.

4.3 Projective Space: A Motivation

We begin this section with the example of two parallel lines,

$$x - 1 = 0 \quad \text{and} \quad x - 2 = 0.$$

These do not intersect in $\mathbb{C}^2$. Suppose we introduced new variables X, Y, Z such that

$$x = \frac{X}{Z} \quad , \quad y = \frac{Y}{Z}.$$

Then our system of equations becomes

$$\frac{X}{Z} - 1 = 0 \;\; m \;\; \frac{X}{Z} - 2 = 0.$$

$$X - Z = \;\; m \;\; X - 2Z = 0.$$

Solving this system, we see that we have solutions $(X, Y, Z) = (0, \lambda, 0)$ for all λ. Translating back to x and y we have $x = 0/0$, $y = \lambda/0$, which would agree with our idea of infinity.

In this, we have basically introduced a new variable, and with this we must be careful. For example, the pair $(X, Z) = (1, 2)$ and $(2, 4)$ both correspond to the same $x = 1/2$. This suggests that what is important are not X and Z, but the fraction X/Z. Less "out of nowhere" than Reid, we thus define the projective plane.

Definition 24. *Let k be a field. Then we define the projective plane*

$$\mathbb{P}^2_k = \{\text{ratios } X : Y : Z \mid X, Y, Z \in k\}$$

with not all three zero.

This may seem a bit fuzzy at the moment, so we will break things up into simpler cases. Suppose that $Z \neq 0$. Then we can "divide by Z" and note that the ratio $X : Y : Z$ is the same as $X/Z : Y/Z : 1$, or $x : y : 1$. Each choice for x and y will yield a different ratio, so when $Z \neq 0$, we can think of the set of all ratios as being k^2, or $\mathbb{A}^2_k$. If $Z = 0$ and $y \neq 0$, then we can do the same thing to get the set of all $x : 1 : 0$ which we can identify with $k = \mathbb{A}^1_k$. Finally, if $Z = Y = 0$, then the only ratio we have is $1 : 0 : 0$, a single point. In conclusion we have

$$\mathbb{P}^2_k = \mathbb{A}^2_k \cup \mathbb{A}^1_k \cup \{\text{point}\}.$$

Our original motivation for introducing the projective plane was to try to make sense of a point at infinity. The above shows that $\mathbb{P}^2$ is just $\mathbb{A}^2$ with an extra copy of k and an extra point. If these are all "points at infinity", why are there so many of them? The geometric explanation is simple: designate one point at infinity for each slope. Thus our copy of $k \cap \{\text{point}\}$ represents all possible slopes (real numbers and infinite), so our projective plane does the trick.

Remark: We try to give philosophical reasons for allowing one point at infinity for each slope of lines. We'd like to have parallel lines intersect at infinity. At the same time we do not want to violate the axiom from Euclidean geometry that lines should intersect in only one point (a super baby case of Bézout's Theorem .) Thus for any given set of parallel lines we want just a single point where they are all to intersect. Now throw another line of a different slope to the mix. It intersects each of these lines in $\mathbb{A}^2$ already, so we do not want it to intersect anything at infinity. Thus the point at infinity for the new line must be a "new" point. Continuing this, you see that we need a point at infinity for each slope value.

4.4 Projective n-space

In this section we give a different definition of the projective plane and in doing so give a better way of generalizing to n dimensions than dealing with ratios.

Definition 25. *Let k be a field. Then we define projective n-space $\mathbb{P}^n_k$ to be the set of lines through the origin in $\mathbb{A}^{n+1}_k$.*

Under this definition, the real projective plane $\mathbb{P}^2_{\mathbb{R}}$ would be the set of all lines through the origin in $\mathbb{R}^3$.

We first note that we can identify the set of lines in $\mathbb{R}^3$ with the set of points in $\mathbb{R}^3 \setminus \{0\}$ provided that we identify points on the same line through the origin - that is, nonzero multiples of each other. Formally, we can think of the set of lines through the origin in $\mathbb{R}^3$ as the set of equivalence classes of points in $\mathbb{R}^3 \setminus \{0\}$ with the equivalence relation $\sim$ defined by

$$(x, y, z) \sim (u, v, w) \iff (x, y, z) = \lambda(u, v, w)$$

for some scalar λ. In other words, ratios $x : y : z$, so our definition is compatible with the one defined earlier. It is this latter definition that we will use more often. In fact we make the following definition:

Definition 26. *Let $(x_0, \ldots, x_n) \neq 0$ be a point in $\mathbb{R}^{n+1}$. Then by $[x_0, \ldots, x_n]$ we denote the point in $\mathbb{P}^n$ corresponding to the line through the origin and $(x_0, \ldots, x_n)$.*

For an example, in $\mathbb{P}^2_{\mathbb{R}}$, $[1, 2, 4] = [2, 4, 8]$ since they represent the same point. The notation $[x_0, \ldots, x_n]$ can be thought of as the equivalence class of $(x_0, \ldots, x_n)$ under the relation $\sim$ defined above.

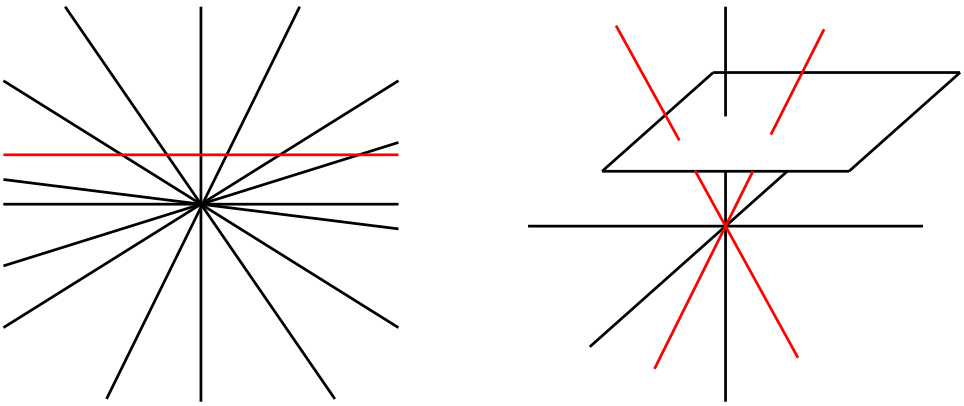


Figure 4: Lines through the origin in $\mathbb{R}^2$ and $\mathbb{R}^3$

We have given a lot of time to $\mathbb{P}^2$ so far, and it's not really fair to $\mathbb{P}^1$. We remedy this now. First note that $\mathbb{P}^0$ is the set of lines through the origin in R^1 which is just a single point. $\mathbb{P}^1_{\mathbb{R}}$ is the set of lines through the origin in $\mathbb{R}^2$. Look at Figure 4. Each line through the origin hits $y = 1$ exactly once, except for the line $y = 0$. Thus we can identify $\mathbb{P}^1_{\mathbb{R}}$ as a line with an extra point,

$$\mathbb{P}^1_R = \mathbb{R} \cup \{\text{point}\} = \mathbb{R} \cup \mathbb{P}^0_{\mathbb{R}}.$$

The last part of this equation is not just for fun. We will see this generalizes.

Consider $\mathbb{P}^2_{\mathbb{R}}$, the set of lines through the origin in $\mathbb{R}^3$. As in the case for $\mathbb{P}^1$ we can consider the plane $z = 1$. Almost every line through the origin in $\mathbb{R}^3$ will meet this plane in exactly one point, that is, all lines except those in the xy plane. But the set of lines in the xy plane is just $\mathbb{P}^1_{\mathbb{R}}$, so we get

$$\mathbb{P}^2_{\mathbb{R}} = \mathbb{R}^2 \cup \mathbb{P}^1_{\mathbb{R}}.$$

This generalizes nicely and we get the following

Proposition 19. *Let k be a field, then*

$$\mathbb{P}^n_k = \mathbb{A}^n_k \cup \mathbb{P}^{n-1}_k$$

4.5 Working with Projective Space

We'd now like to start discussing how we can fit polynomials into the frame of projective space. From now on, if the field is implied or unimportant then we may omit the subscript and just write $\mathbb{P}^n$. We saw before that we could take an equation in $\mathbb{R}[x, y]$ and turn it into an equation of $\mathbb{R}[X, Y, Z]$, by making the substitution

$$x = \frac{X}{Z} \quad , \quad y = \frac{Y}{Z}.$$

This process is called *homogenization* and can be done in general.

Definition 27. *Let R be a ring and let f be a polynomial in $R[x_1, \ldots, x_n]$. The corresponding homogeneous polynomial for f is the polynomial $f(\frac{X_1}{X_0}, \ldots, \frac{X_n}{X_0})$ in $R[X_0, X_1, \ldots, X_n]$. (If we have less than 4 variables, it's customary to use X, Y, Z instead of subscripts.)*

Example: We homogenize equations from $\mathbb{R}[x, y]$ into $\mathbb{R}[X, Y, Z]$:

- $x^2 + y^2 = 1$ goes to $(X/Z)^2 + (Y/Z)^2 = 1$ or $X^2 + Y^2 = Z^2$.
- $y = x^2$ goes to $Y/Z = (X/Z)^2$ or $YZ = X^2$.

A nice shortcut for doing this is to search for the highest degree term, and the multiply each lower degree term by Z (or x_0) to bring everything to the same degree.

Definition 28. *A polynomial f in $R[x_0, \ldots, x_n]$ is called homogenous polynomial of degree d if every term has degree d. Such a polynomial is called a form of degree d. The set of all forms of degree d is denoted R_d. (Here the number of variables, is understood).*

Homework 33. *Prove that $f(x_0, \ldots, x_n)$ is a homogenous polynomial of degree d if and only if $f(\lambda x_0, \ldots, \lambda x_n) = \lambda^d f(x_0, \ldots, x_n)$, for all scalars λ.*

It is this property that we will find very useful in our study of homogeneous polynomials.

You'll recall that in the first section of these notes we defined a map V so that $V(f)$ is the zero locus of f in $\mathbb{A}^n_{\mathbb{R}}$. In $\mathbb{A}^n_{\mathbb{R}}$, f could be any polynomial. If we try to extend this property to projective space, we run into trouble. We illustrate this problem in the next example.

Example: Consider the polynomial $f(X, Y, Z) = XY + Z^3$. Then we can evaluate f at a point in $\mathbb{P}^2_{\mathbb{R}}$. But be careful! $f([0, 0, 1]) = 1$ while $f([0, 0, 2]) = 8$ but in $\mathbb{P}^2_{\mathbb{R}}$, $[0, 0, 1] = [0, 0, 2]$, so "evaluation" isn't well defined.

Our luck isn't much better if we only look at the zero-locus. For example, $f([2, 4, -2]) = 0$, but $f([1, 2, -1]) \neq 0$. Luckily, in special cases, the zero locus causes no problems, so that we can make sense of $V(f)$. We handle this in the following proposition.

Proposition 20. *Let f be a homogeneous polynomial of some degree d. Then the vanishing locus of f, $V(f)$ is well-defined in $\mathbb{P}^2_{\mathbb{R}}$.*

Proof. Suppose that a is some point in $\mathbb{P}^2_{\mathbb{R}}$ such that $f(a) = 0$. Then $\lambda^d f(a) = 0$ for each scalar λ and thus by earlier remarks, $f(\lambda(a)) = 0$. Thus f actually vanishes at $[a]$ and the zero locus is well-defined. □

Remark: We repeat here an important hint: It is often useful to think of $\mathbb{P}^2_{\mathbb{R}}$ as just $\mathbb{R}^3 \setminus \{0\}$ with multiples of vectors being considered the same.

We now have a map

$$V : \{\text{homogenous polynomials of } R[x_0, \ldots, x_n]\} \longrightarrow \{\text{subsets of } \mathbb{P}^n_R\}\,.$$

A map I going the other direction can also be defined. For us the idea of homogeneous ideals is what is really important, so we will not discuss this development.

4.6 Bézout's Theorem Revisited

You'll recall that our whole motivation for introducing projective space was to make Bézout's Theorem more precise; to better count the number of intersection points of curves. Our hard work pays off with the following theorem.

Proposition 21. *(Bézout's Theorem for the Projective Plane) Let k be an algebraically closed field. Let C and D be two curves in $\mathbb{P}^2_k$ with no common component of degrees c and d respectively. Then counting multiplicity, C and D meet in exactly cd points in $\mathbb{P}^2_k$.*

We will actually prove this later on in the course of these notes using projective resolutions. For now we show the power of this theorem and give examples.

Examples: Consider the two parabolas $y = x^2 + 1$, $y = -x^2$. They do not meet at all in $\mathbb{R}^2$, but Bézout's Theorem tells us that they will meet in 4 points in $\mathbb{P}^2_{\mathbb{C}}$. In the complex plane $\mathbb{C}^2$ they meet in the two points $(\pm i\sqrt{2}/2, 1/2)$. We hope that the point at infinity has multiplicity two to bring us up to four. Going to homogeneous coordinates our equations become

$$YZ = X^2 + Z^2 \quad , \quad YZ = -X^2.$$

Solving, we get $2X^2 + Z^2 = 0$. If $Z = 0$ (our point at infinity) then $2X^2 = 0$ so we get $X = 0$ of multiplicity two as required.

Homework 34. *The parallel parabolas $y = x^2$ and $y = x^2 + 1$ meet at infinity. How can you count four points of intersection in the projective plane?*

5 Linear Algebra and Homogeneous Polynomials

In the previous section we defined the set R_d the set of forms of degree d in $R = \mathbb{R}[x_0, \ldots, x_n]$. Note that this is not a ring for many reasons (find at least two for homework). $R_d \cup \{0\}$ is a vector space over $\mathbb{R}$, however. Indeed, adding

two forms of degree d is another form of degree d, and similarly for scalar multiplication.

Now that we have a vector space, we recall that every vector space has a basis. So what is a basis of R_0? A form of degree 0 is just a constant, so $\{1\}$ is a basis. A basis for forms of degree 1 is $\{x_0, \ldots, x_n\}$, so R_1 has dimension $n+1$. To make the notation easier for higher degrees, we simplify to three variables, $\mathbb{R}(x, y, z)$. $(n = 2)$

Degree	Basis	Dimension of R_d
1	$\{x, y, z\}$	3
2	$\{x^2, y^2, z^2, xy, yz, xz\}$	6
3	$\{x^3, y^3, z^3, x^2y, x^2z, xy^2, y^2z, xz^2, yz^2\}$	10

The pattern of triangular numbers suggests that the dimension of R_d here is $\binom{d+2}{2}$. In fact we have the following:

Proposition 22. *The dimension of R_d in $k[x_0, \ldots, x_n]$ is*

$$\binom{d+n}{n.}$$

Proof. Clearly a basis for R_d is the set

$$S = \{x_0^{a_0} \cdots x_n^{a_n} \mid a_0 + \ldots + a_n = d, a_i \geq 0\}.$$

The trick will be in counting how many elements it has. This can be done in the following way. Consider $n + d$ bowls in a line. Fill any n with water. From this we construct an element of S as follows. Let a_0 be the number of bowls to the left of the first filled bowl. Let a_1 be the number of bowls between the first and second filled bowls, etc. Since there are exactly n dry boxes, $a_0 + \cdots + a_n = d$. For example, we show how to get the polynomial xy^2z in $R[x, y, z]$ in Figure 5.

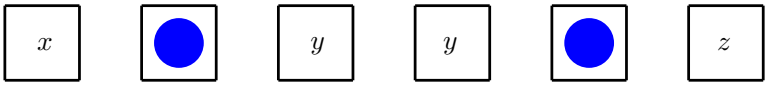


Figure 5: Choosing 2 bowls from 5 to get xy^2z

A bit of thought shows that any polynomial in S can be gotten this way. Since we are free to choose n bowls from our $n + d$ the result follows. □

So we now have a sequence of vector spaces $R_0, R_1, R_2, \ldots$ and we know their dimensions. This leads to a very natural question

Question: What is interesting about the subspaces of the R_i?

We motivate this with the following proposition.

Proposition 23. *Let X be a subset in $\mathbb{P}^n$. Then the set of homogeneous polynomials of degree d that vanish on X form a vector space.*

Proof. Let $f, g \in R_d$ and f and g vanish on X. Thus for every point $[a] \in X$, $f([a]) = g([a]) = 0$. So clearly $f+g$ vanishes on X as does every scalar multiple of f. Since we're only dealing with homogeneous polynomials, we know that the vanishing is well-defined from previous discussion. □

We will denote the set of homogeneous polynomials of degree d that vanish on X as $(I_X)_d$. For example, if $X = \mathbb{P}^n$, then $(I_X)_d = 0$ since the zero polynomial is the only polynomial vanishing at every point. If $X = \emptyset$, then $(I_X)_d = R_d$ for each d since every polynomials vanishing on the empty set. (It's quite silly really.)

Our goal in this section will be to discuss the vector subspace of homogeneous polynomials vanishing on a set X. For us, it will be a great achievement if we can determine the dimension of such a subspace. To do so for an arbitrary set X is a bit unwieldy, so we begin with the simple case when X is a set of points in $\mathbb{P}^2$.

5.1 Conditions Determined by Points in $\mathbb{P}^2$.

Suppose we are given a $Z = \{p_1, \ldots, p_r\}$, a set of r points in $\mathbb{P}^2$. We would like to determine all forms of degree d that vanish on these points. Let $f \in R_d$ be a form of degree d. Then since R_d has dimension $m = \binom{d+2}{2}$, we have a basis of R_d, $\{e_1, \ldots, e_m\}$. Thus

$$f = a_1 e_1 + \ldots + a_m e_m$$

for some constants $a_1, \ldots, a_m \in R$. But if f is to vanish at p_1, then we can plug in the point p_1 into the above equation (remember that the e_i are just monomials of degree d) and get some equation

$$c_{11}a_1 + \ldots + c_{1m}a_m = 0.$$

We can actually do this for each point p_i so we in fact get a system of equations

$$\begin{array}{ccccccc} c_{11}a_1 & + & \ldots & + & c_{1m}a_m & = & 0 \\ c_{21}a_1 & + & \ldots & + & c_{2m}a_m & = & 0 \\ & & & \ldots & & & \\ c_{r1}a_1 & + & \ldots & + & c_{1m}a_m & = & 0. \end{array}$$

Remember, our goal is to find the constants $a_1, \ldots, a_m$ so this is now just a question of linear algebra. We have m unknowns and r equations. A nice intuitive way to think about this is that to begin with, your homogeneous polynomial is free to be whatever it wishes if it has no vanishing restrictions. But as soon as the function is required to vanish at a certain point, that places one restriction on its behavior. It "loses a degree of freedom" if you will. Either from this viewpoint, or from linear algebra and knowledge of systems with m unknowns and r equations, we have the following.

Proposition 24. *Let Z be a set of r points in $\mathbb{P}^2$. Then $dim(I_Z)_d \geq R_d - r$.*

Proof. Sketch: Each point imposes conditions on the set $(I_Z)_d$. If every point takes away one degree of freedom, then we will lose r dimensions. Of course, the conditions imposed by the points might very well be redundant, so we don't necessarily have equality. □

We will now do several examples to familiarize you with the abstraction above so that you feel more at ease. Suppose we are looking for polynomials that vanish on the set $Z = \{[1, 0, 1], [0, 1, 1]\}$. Clearly no nonzero form of degree 0 can vanish on both of these points, so $\dim(I_Z)_0 = 0$. (You will find this is almost always the case, in fact you could probably prove that now!) Next we try to tackle forms of degree 1. Forms of degree 1 look like

$$f = AX + BY + CZ.$$

If they vanish at $[1, 0, 1]$ and $[0, 1, 1]$, then we get the system of equations

$$A + C = 0$$

$$B + C = 0.$$

Solving this system, we see that $A = -C$, $B = -C$. Thus all solutions are of the form

$$(A, B, C) = (-C, -C, C) = C(-1, -1, 1)$$

which is one dimensional. In this case, we started with 3 degrees of freedom, but the two points knocked us down to 1. This should also make intuitive sense, since a form of degree 1 is just a line, and given two points, there is a unique line passing through them. The linear systems approach begins to get a bit ridiculous even at the next stage. We will write down the system, but not mention the solution. Better methods will be introduced shortly which will make these calculations unnecessary. A form of degree 2 looks like

$$f = AX^2 + BY^2 + CZ^2 + DXY + EXZ + FYZ.$$

Plugging in our two points yields the system

$$A + C + E = 0$$

$$B + C + F = 0.$$

The solution to this will have dimension 4 as it will have 4 free variables. Note that we could have also written the system in matrix form

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} A \\ B \\ C \\ D \\ E \\ F \end{pmatrix} = 0.$$

Since the matrix has rank 2, we expect the solution space to be 6-2=4 dimensional.

We next do an example where equality does not hold in Proposition 24. Consider the set $Z = \{[1,0,1],[0,0,1],[-1,0,1]\}$. Then these three points lie in a line. We compute $\dim(I_Z)_1$. Let $f \in R_1$. Then

$$f = AX + BY + CZ.$$

The points give us the system

$$A + C = 0$$

$$C = 0$$

$$-A + C = 0$$

which has as a solution all triples $(A, B, C) = (0, B, 0)$, a one dimensional vector space. In this case, our 3 points did not each knock off a dimension, only 2 of them did. The geometric reason for this is that if we are looking for a line vanishing at the first two points, this uniquely determines the line. The third point just happens to fall on that same line so it doesn't add any new restrictions. If that point were not on the line (say the last point of Z was $[1,1,0]$) then you should check that there are no forms of degree 1 vanishing on Z. Thus the dimension would be $3 - 3 = 0$.

Before we do another example we now note that as these examples indicate, the linear algebra might involve several equations and it may be a mess. To remedy this, we develop a theory to help us with our work. We will also clarify the different between the first two examples. We start with a definition.

Definition 29. *Let Z be a set of r points in $\mathbb{P}^2$. We say that Z imposes independent conditions on forms of degree d if*

$$dim(I_Z)_d = \binom{d+2}{2} - r,$$

that is, if equality holds in Proposition 24.

In other words, a set imposes independent conditions if every one of the points knocks the degree down. In the two examples above, the first set imposed independent conditions on forms of degree one, since $\dim(I_Z)_d = 3 - 2 = 1$, but the second did not because there were three points yet the dimension was still 1.

When a set imposes independent conditions on forms of degree d this is extremely nice for our calculations since it immediately gives us the dimension of $(I_Z)_d$. The trouble is that at the moment this is actually how we are defining independent conditions. The next proposition will give a new nice criterion for determining if a set imposes independent conditions and will be one we use quite frequently.

Proposition 25. *Let Z be a set of r points in $\mathbb{P}^2$. Then Z imposes independent conditions on forms of degree d if and only if for each point $p \in Z$, there exists a form of degree d vanishing on the other $r-1$ points, but not vanishing at p.*

Proof. Suppose first that Z imposes independent conditions on forms of degree d. Let $m = \dim(R_d) = \binom{d+2}{2}$ and let $\{e_1, \ldots, e_m\}$ be a basis for R_d. Then if f vanishes on Z then

$$f = a_1 e_1 + \cdots + a_m e_m$$

for some constants $a_i \in R$. Plugging in the coordinates of each of the r points of Z we get a matrix as before

$$\begin{pmatrix} c_{11} & \cdots & c_{1m} \\ c_{21} & \cdots & c_{2m} \\ & \cdots & \\ c_{r1} & \cdots & c_{rm} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix} = 0.$$

Call the rows of this matrix $b_1, \ldots, b_r$. Since the points impose independent conditions, the dimension of the nullspace of this matrix is $m - r$, so it follows that the b_i are all linearly independent. Let our "coefficient vector" be $v = (a_1, \ldots, a_m)$. Then the matrix equation tells us that $b_i \cdot v = 0$ for each i. In other words, each row is perpendicular to v. We now pause to collect our thoughts and think about what we are trying to prove.

Goal: Pick any point $p \in Z$. We are trying to find a form f of degree d that vanishes at each point in Z except p.

Let us translate this into our notation. "Picking a point p" is like picking a row of our matrix; "Finding a form f" is the same as finding $a_1, \ldots, a_m$; "Vanishing at each point in Z except p" means that $(a_1, \ldots, a_m)$ is perpendicular to every row of our matrix except the one we selected. Thus we can reformulate our problem as follows

Restatement: Let V be a vector space of dimension m, Then for any set of r linearly independent vectors ($r < m$) $v_1, \ldots, v_r$, we can find a vector w such that $w \cdot v_i = 0$ for each $i = 1, \ldots, r-1$, but $w \cdot v_r \neq 0$.

The proof of this is simple: Consider the space $U = \mathrm{Span}(v_1, \ldots, v_{r-1})$. This is an $r-1$ dimensional space. So $U^\perp$, the set of all vectors perpendicular to U is an $m - (r-1)$ dimensional space. We may assume $r < m$. (proof for $r = m$ is left to the reader) Then $U^\perp$ is at least a two dimensional vector space. Since $(v_1, \ldots, v_r)$ is independent, v_r is not in U so we can choose w in $U^\perp$ not perpendicular to v_r. Then by construction, w is perpendicular to $v_1, \ldots, v_{r-1}$ as required.

Conversely, suppose that we can remove any point p from Z and there exists a form of degree d vanishing on all points of Z but not p. Then using the notation above, this means for each i there exists a vector v_i such that

$$v_i \perp b_1, b_2, b_{i-1}, b_{i+1}, \ldots, b_r,$$

and $v_i \cdot b_i \neq 0$. Then we claim that $b_1, \ldots, b_r$ are linearly independent. Suppose that $c_1 b_1 + \ldots c_r b_r = 0$. Then dotting both sides with v_i shows that $c_i = 0$, so

the b_i are linearly independent and thus the matrix has rank r meaning that the dimension of $(I_Z)_d$ is $m - r$ as required. □

The previous proof can certainly be improved. Besides the Nullstellensatz, this is the first lengthy argument we've had to do. What is important in this definition is the statement, and not the proof. To illustrate its use, we give some examples. We will draw all the sets of points in the affine plane (it's impossible to do otherwise), but it is understood they are in $\mathbb{P}^2$.

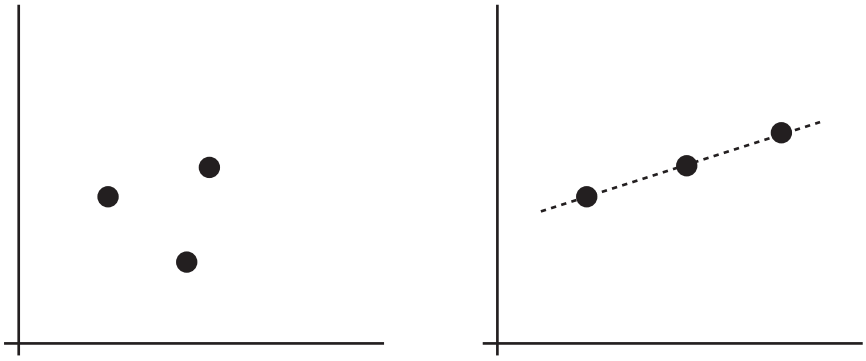


Figure 6: Two sets of points; only the first imposes independent conditions on forms of degree one.

Example: We first consider the two sets in Figure 6. The first set imposes independent conditions on forms of degree 1. To see this, remove any point. Then there is a line (a form of degree 1) passing through the other two but not the point you removed. Thus the first set imposes independent conditions on forms of degree one.

The second set does not impose independent conditions on forms of degree 1 because to draw a line through two points, the line will also contain the third.

We make the note that both sets impose independent conditions on forms of degree 2. To see this, just remove a point and draw a line through each of the remaining points so that it does not intersect the third. The union of these two lines is a conic.

Example: Consider the two sets in Figure 7. Neither imposes independent conditions on forms of degree 1 (this is clear), and from Figure 8 it is clear that the second set imposes independent conditions on forms of degree two.

To see that the first set does not, we will have to use Bézout's Theorem. Indeed, remove one of the four points lying on the line. Suppose that C is some form of degree two passing through the other 4 points. Then it must pass through the other 3 points on the line. But this would imply that the form of degree 2 meets the line in 3 points. By Bézout's Theorem this means that the conic must have a line as a component, and thus we cannot avoid the point we

removed. Finally, it is easy to see that both sets in Figure 7 impose independent conditions on forms of degree 3.

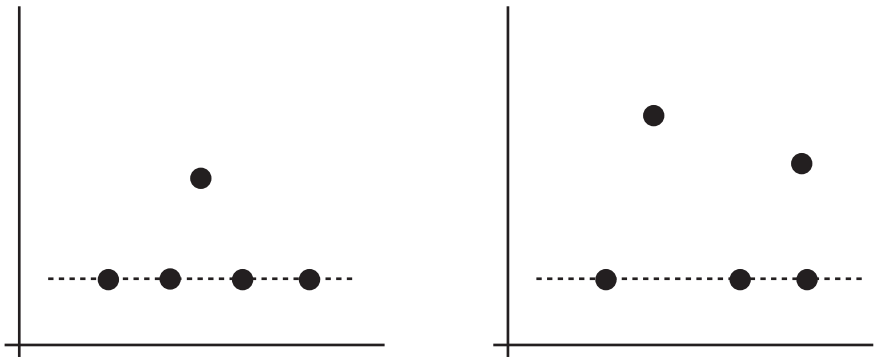


Figure 7: Two sets of points; Neither imposes independent conditions on forms of degree one, and only the second imposes independent conditions on forms of degree two

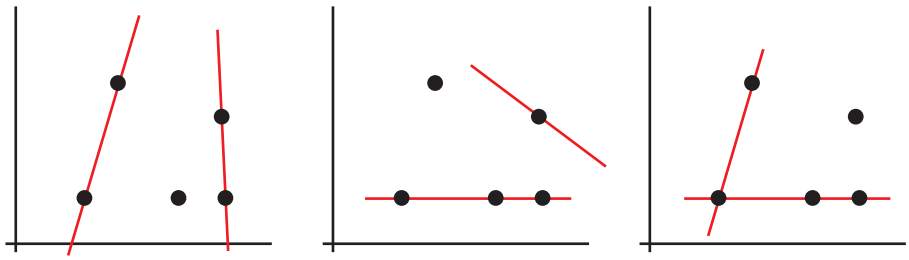


Figure 8: Some illustrations showing that the second set in Figure 7 imposes independent conditions on forms of degree two

In these two examples we came across a very useful fact which we state in the next proposition.

Proposition 26. *Let Z be a set of points in $\mathbb{P}^2$ that imposes independent conditions on forms of degree d. Then Z also imposes independent conditions on forms of degree $d+1$.*

Remark: This proposition along with induction shows that once a set starts

imposing independent conditions, it does so thereafter. Thus once we get independent conditions, we know how to compute $\dim(I_Z)_d$ thereafter.

Proof. To show that Z imposes independent conditions on forms of degree $d+1$, pick any point $p \in Z$. Since Z imposes independent conditions on forms of degree d, we know there exists a form of degree d (call it C) that passes through all of Z except the point P. Now just take a line L not passing through p and the union $L \cup C$ is a form of degree $d+1$ not passing through p as required. □

We end this section by reviewing the progress that we have made on computing $\dim(I_Z)_d$. Originally our only method was to solve a system of equations and look at the dimension of the vector space. We next found a nice geometric condition that tells us when a set of points imposes independent conditions. Finally, we found that once a set imposes independent conditions, it does so henceforth.

6 Hilbert Functions of Points and Ideals

In this section we plan to compute what is known as the Hilbert function of a set of points, and then in general we will define what we mean by the Hilbert function of an ideal. We begin with an example from the previous section.

Let Z be a set of 5 points in $\mathbb{P}^2$, four of them collinear. (e.g. the situation in the first set of Figure 7. We will compute $s(d) = \dim(I_Z)_d$ for a few values of d.

If $d = 0$, then $s(d) = 0$ as usual. Since the points are not all collinear, $s(1) = 0$ as well. The computation of $s(2)$ is much more interesting. Let C be any conic vanishing on Z. Then by Bézout's Theorem we know it must contain the line L through the 4 points as a factor. To get the other factor, we can multiply L by any line passing through the 5th point. Lines are of the form $AX + BY + CZ$ and the one point will clearly impose one condition so we still have two degrees of freedom. Thus, $s(2) = 2$. Finally, we know that Z imposes independent conditions on forms of degree 3 and higher, so $s(d) = \binom{d+2}{2} - 5$ for all $d \geq 3$. Thus we can make the following table:

degree (d)	0	1	2	3	4	5	6	7
$\dim(R_d)$	1	3	6	10	15	21	28	36
$s(d)$	0	0	2	5	10	16	23	31
$\dim(R_d) - s(d)$	1	3	4	5	5	5	5	5

The bottom row of the table represents the difference between the dimension of R_d and the dimension of the subspace vanishing on Z. Thus it represents how much smaller $(I_Z)_d$ is than R_d. This difference is very important and it is called the *Hilbert Function*.

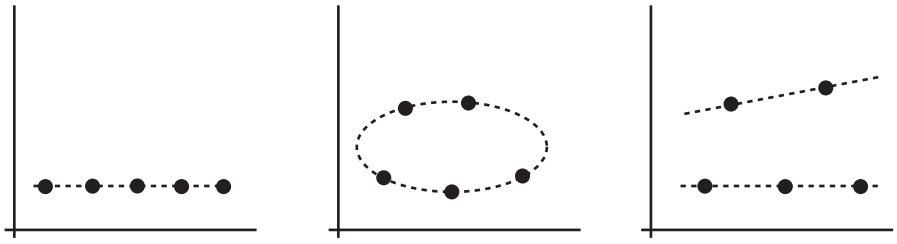


Figure 9: Some various arrangements of 5 points

Definition 30. *Let Z be a finite set of points in $\mathbb{P}^n$. The Hilbert Function of Z is the sequence $\{H_Z(i)\}_{i=0}^{\infty}$ where*

$$H_Z(d) = dim(R_d) - dim(I_Z)_d = \binom{d+n}{n} - dim(I_Z)_d.$$

Homework 35. *Compute the Hilbert function of for each of the sets of points in Figure 9.*

Homework 36. *Compute the Hilbert function of d collinear points in $\mathbb{P}^2$.*

In general, computing the Hilbert function of a set of points is highly nontrivial. The function is extremely dependent on the geometry of the set of points. For instance, if the points are all collinear, it is easy to compute the function. In general, this is not the case. To illustrate why this is not the case consider the following. Suppose you had 9 points in $\mathbb{P}^2$ and you were wondering if they imposed independent conditions on forms of degree 3. Thus you ask if there is a cubic passing through any 8 of them, but not the ninth? Just glancing at the points, you cannot say as we have the following classical theorem.

Proposition 27. *(Cayley Bacharach) Let C and D be two cubic curves in $\mathbb{P}^2$ intersecting in 9 points. (As in Figure 10.) Then any cubic curve passing through 8 of these points passes through the ninth as well.*

Thus it is not so easy to tell when a set of points imposes independent conditions. For nine points the answer is no if they lie on two distinct cubic curves! One should not abandon hope, however. Although it may be very difficult to compute the entire Hilbert Function, as you may have noticed in the homework computations, eventually the function levels off and remains constant. In fact it is constantly equal to the number of points!

Proposition 28. *Let Z be a set of r points in $\mathbb{P}^2$. Then $H_Z(d) = r$ for all $d \geq r - 1$. Furthermore, this is the best lower bound we can achieve.*

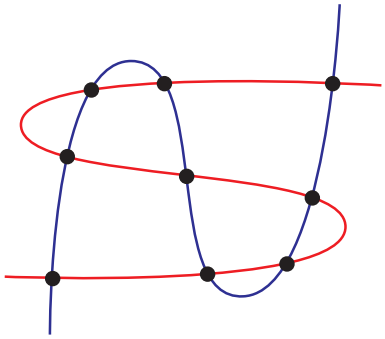


Figure 10: Two cubics intersecting in 9 points

Proof. We note that Z imposes independent conditions on forms of degree $r-1$ since removing any one point p, we can just draw lines through the remaining $r-1$ point, carefully avoiding p. Their union is a form of degree $r-1$. To see this is the best lower bound, note that the Hilbert Function of r collinear points is $H_Z(d) = d+1$ if $d < r$ and r afterwards. □

This theorem has a beautiful application, using it we will be able to give a two line proof of Bézout's Theorem. We will later find a convenient way to compute the Hilbert function of the intersection of two curves. Since this will be a set of points, we can simply see where it levels off and that will give us the number of intersection points. To do this, we will need more algebraic machinery, however. In the meantime, here's another interesting fact about Hilbert functions.

Proposition 29. *If Z is a set of points in $\mathbb{P}^2$ then for any t, $h_z(t) \leq h_z(t+1)$.*