Chapter 10
Geometry of Definable Sets

Abstract In the previous chapter we saw examples of mathematical structures that are simple enough to allow a complete analysis of the parametrically definable subsets of their domains. Those structures are of some interest, but the real objects of study in mathematics are richer structures such as $(\mathbb{Q}, +, \cdot)$ or $(\mathbb{R}, +, \cdot)$. To talk about them we first need to take a closer look into their definable sets. Definable sets in each structure form a geometry in which the operations on sets are unions, intersections, complements, Cartesian products, and projections from higher to lower dimensions. We will see how those operations correspond in a natural way to Boolean connectives and quantifiers, and how the name "geometry" is justified when it is applied to sets definable in the field of real numbers. The last two sections are devoted to a discussion of the negative solution to Hilbert's 10th problem.

Keywords Boolean combinations · Existential quantification and projections · Diophantine equations · Hilbert's 10th problem · Tarski-Seidenberg theorem

10.1 Boolean Combinations

The geometry of definable sets involves two levels. At the first level, Boolean connectives[1] $\wedge$, $\vee$, and $\neg$ are used to combine the basic relations of the structure giving rise to new definable sets. The Boolean connectives correspond closely to basic operations on sets, which are also called *Boolean operations*. Conjunctions of formulas correspond to intersections of sets defined by them, disjunctions correspond to unions, and negations to complements. If A is the domain of a structure $\mathfrak{A}$, then for each natural number n, the definable subsets of A^n form a *Boolean algebra* of sets. Let $\mathfrak{B}$ be set of subsets of some set U. Then $\mathfrak{B}$ is a Boolean algebra, if for all X and Y in $\mathfrak{B}$ their intersection $X \cap Y$, their union $X \cup Y$, and their complements in U are also in $\mathfrak{B}$.

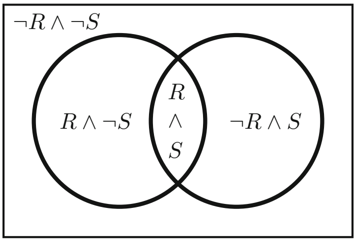
[1] After George Boole (1815–1864).

R. Kossak, *Mathematical Logic*, Springer Graduate Texts in Philosophy 3,
https://doi.org/10.1007/978-3-319-97298-5_10

At the next level, quantifiers are employed to bind free variables in formulas. A formula with n free variables defines a subset of the n-th Cartesian power of the domain. Appending quantifiers results in a formula with fewer free variables that defines a set in a lower Cartesian power. This is an important aspect of the geometry of definable sets and we will discuss it in detail in the next section.

Let $\mathfrak{A}$ be a structure with the domain A and two unary[2] relations R and S. We the use the same characters for relation symbols and the corresponding relations. Combining atomic formulas $R(x)$ and $S(x)$ using connectives $\wedge$, $\vee$, and $\neg$, we can form their *Boolean combinations*, such as $R(x) \wedge S(x)$, or $\neg R(x) \vee S(x)$. In the case of just two relation symbols, there is a helpful diagram that illustrates all possible combinations.

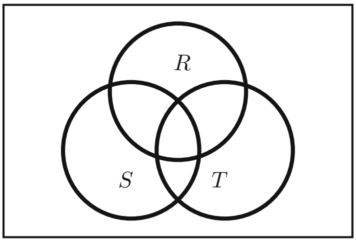
The picture below is the *Venn diagram* for the two sets R and S. The rectangle represents the domain of the structure. The two circles represent the sets R and S, their intersection represents the set defined by the formula $R(x) \wedge S(x)$, and the area outside the two circles is defined by $\neg R(x) \wedge \neg S(x)$. All Boolean combinations of $R(x)$ and $S(x)$ can be represented as regions in the diagram.



How many different Boolean combinations are there? In principle, there are infinitely many, because we can keep combining formulas to obtain longer and longer Boolean combinations, but by inspecting the diagram one can see that the only sets that can be defined by those formulas are combinations of different regions demarcated by the two circles. Cutting the rectangle along the circles will give us four pieces of the diagram. Any collection of those pieces corresponds to a subset of the domain—a unary relation that is definable from R and S. Since any choice of the pieces gives us a definable set, assuming that all pieces represent nonempty sets, altogether it gives us $2^4 = 16$ different definable sets, and that already includes the sets R and S, the empty set (no pieces), and the whole domain (all pieces).

To analyze all possible Boolean combinations one can obtain from three relations, we can inspect a Venn Diagram with three circles. Let us call the relations R, S, and T.

[2] Recall that a unary relation is a subset of the domain of a structure.



The three circles cut the domain into eight pieces. Any collection of those pieces gives us a definable set, so, again assuming they each piece represents a nonempty set, there are $2^8 = 256$ such definable sets. The diagram provides a visualization, and, in a sense, a complete understanding of what those sets are.

For more than three unary relations, no diagram with circles can represent all Boolean combinations, but there is nothing special about circles. One can use other geometric shapes, but the pictures are less transparent. The Wikipedia article on Venn diagrams gives interesting examples. Regardless of visual representations, the diagrams show that in a structure with a finite number of relations, Boolean combinations of the atomic formulas generate a large, but finite set of new definable relations.

In this introduction to the geometry of definable sets, so far we have only considered unary relations. In general, relations can be subsets of any Cartesian power of the domain of the structure. Boolean combinations of n-dimensional relations are definable subsets of the n-th Cartesian power of the domain. Why we call this geometry will become clearer in the next section in which we will see how logic allows us to define new sets in all possible dimensions, and how those sets interact with one another.

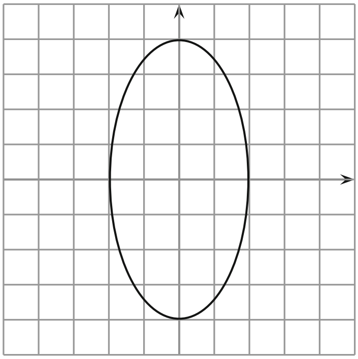
10.2 Higher Dimensions

Let us go back to the structure $\mathfrak{A}$ with domain A and two unary relations R and S. To define new relations other than the Boolean combinations of the basic relations, we will now take advantage of the infinite supply of variables. According to Definition 7.2, atomic formulas $R(x)$ and $R(y)$ define the same subset of A, namely the set R. Boolean combination $R(x) \wedge R(x)$, also defines R. Saying $R(x)$ twice does not make the statement any stronger. The situation changes when we consider $R(x) \wedge R(y)$. This formula has two free variables, hence it defines a subset of the Cartesian square A^2. It defines a set in two dimensions. Similarly, the formula $R(x) \wedge R(y) \wedge R(z)$ defines a subset of A^3, and in general the formula $R(x_1) \wedge R(x_2) \wedge \cdots \wedge R(x_n)$, with n free variables, defines a subset of A^n.

One could argue that the relation defined by $R(x) \wedge R(y)$ does not reveal any new features of the structure. Its "information content" is not much different than that of the set R itself. But a more complex picture emerges when we consider the relations such as $R(x) \vee R(y)$, and $\neg(R(x) \vee R(y))$, $R(x) \wedge S(y)$, and many other that can be obtained by forming Boolean combinations of atomic formulas with different choices of variables. A whole rich architecture of higher dimensional relations emerges, and the full analysis of the structure $\mathfrak{A}$ must include some insight into what this architecture is.

We have not defined dimension formally, and we will not do that. Informally, one could refer to subsets of the n-th Cartesian power A^n as n-dimensional sets, but this is not precise. A straight line as a subset of a plane, or the three-dimensional space, is still a one dimensional object. The n-th Cartesian power A^n has subsets of all dimensions up to n.

So far we have only considered definable sets that are Boolean combinations of the basic relations of the structure. It is time to take a look at the role of quantifiers. We will start with an example. In the field of real numbers $\mathfrak{R}$ the formula $4x^2 + y^2 - 16 = 0$ defines a subset E of $\mathbb{R}^2$ illustrated by the picture below.



The set E is an ellipse with the minor axis of length 4 and the major axis of length 8. Let $\varphi(x)$ be the formula $\exists y[4x^2 + y^2 - 16 = 0]$, and let $\psi(y)$ be the formula $\exists x[4x^2 + y^2 - 16 = 0]$. Each of these formulas defines a set of real numbers. The first defines the interval of all numbers between -2 and 2, including the endpoints. The second defines the interval between -4 and 4, also with the endpoints. Those two sets are images under *projections* of the set E onto the horizontal and the vertical axes. Binding free variables in a formula by existential quantifiers results in definitions of images under projections of definable sets in higher dimensions onto lower dimensions. Projections are among the most important geometric operations; hence, this feature of existential quantification provides a strong link between geometry and logic. Let us examine this phenomenon in greater detail.

A quadratic equation in two variables is an equation of the form

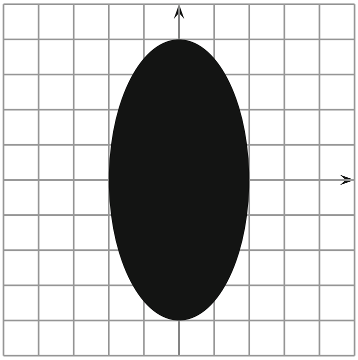
$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \tag{$*$}$$

where x and y are variables representing the unknowns, and a, b, c, d, e, and f represent parameters. In our example above $a = 4$, $c = 1$, $f = -16$, and all other parameters are equal to 0. Each equation of the form $(*)$, in which the parameters are real numbers, is a formula of the language of $\overline{\mathfrak{R}}$, and it defines a subset of $\mathbb{R}^2$. Such sets are known as *conic sections*. Each conic section is either a circle, an ellipse, a parabola, a hyperbola, or, when both a and c are 0, a straight line, or two straight lines, as, for example, is the case for the equation $x^2 - y^2 = 0$.[3] For some choice of parameters, the corresponding conic section is empty. For example, it is empty for the equation $x^2 + y^2 + 1 = 0$ which has no real solutions.

Since the order relation $<$ is definable in $\mathfrak{R}$, the solution sets to inequalities of the form

$$ax^2 + bxy + cy^2 + dx + ey + f < 0, \qquad (**)$$

are also definable in $\overline{\mathfrak{R}}$. The formula $4x^2 + y^2 - 16 < 0$ defines the interior of the ellipse E.



Boolean combinations of the solution sets to equations $(*)$ and inequalities $(**)$ are also definable in $\overline{\mathfrak{R}}$. We already see some interesting geometry, and it is only based on polynomial equations in two variables and of degree 2. A much richer picture emerges when we also consider polynomials of higher degrees, and equations and inequalities with more variables. It becomes even more interesting when we also include projections from higher to lower dimension, but before we see how, we need a few more words of explanation why we call all this a geometry.

10.2.1 Euclidean Spaces

The Cartesian powers $\mathbb{R}^n$ serve as models of classical geometry. $\mathbb{R}^1$ is just the one-dimensional line $\mathbb{R}$; $\mathbb{R}^2$ is the two-dimensional plane; $\mathbb{R}^3$ is the three-dimensional space; and so on. Points in $\mathbb{R}^2$ are ordered pairs of numbers, points in $\mathbb{R}^3$ are ordered triples, and in general a point in $\mathbb{R}^n$ is a sequence of n real numbers. Each space $\mathbb{R}^n$

[3] Notice that another special case $ax^2 + by^2 = 0$ the defined set is just $\{(0, 0)\}$. One point can be considered as a circle that has radius 0.

is equipped with a metric, that is a function computing distances between points. The distance between two points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ in $\mathbb{R}^n$ is

$$\sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

For example, the distance between $x_1 = 2$ and $y_1 = 5$ in $\mathbb{R}^1$ is $\sqrt{(2-5)^2} = \sqrt{9} = 3$.[4] The distance between $(x_1, x_2) = (0, 0)$ and $(y_1, y_2) = (1, 1)$ in $\mathbb{R}^2$ is $\sqrt{(1-0)^2 + (1-0)^2} = \sqrt{2}$.

The set $\mathbb{R}^n$ with the distance function defined above is known in mathematics as an *Euclidean metric space*. In general, a *metric space* is a set M equipped with a distance function d, defined on M^2 with values in $\mathbb{R}$, which satisfies the following conditions for all x, y, and z in M:

- $d(x, y) = 0$ if and only if $x = y$. The distance between the point and itself is 0, and the distance between two different points is never 0.
- $d(x, y) = d(y, x)$. It is as far from x to y, as is from y to x.
- $d(x, z) \leq d(x, y) + d(y, z)$. This is the *triangle inequality*. It says that it is never shorter to go first from x to y, and then from y to z, than directly from x to z.

The distance formula for $\mathbb{R}^n$ clearly satisfies the first two conditions, and it can be shown that it also satisfies the third.

Metric spaces are classical mathematical structures, but they do not fit our definition of structure. The reason is that the relation given by the distance function is between pairs of points in M and real numbers, and real numbers are usually not elements of M. There is a way to convert each metric space into a first-order structure by considering a structure with the domain $M \times \mathbb{R}$, but in the case of the Euclidean spaces $\mathbb{R}^n$, $\mathbb{R}^n \times \mathbb{R}$ can be identified with $\mathbb{R}^{n+1}$, so there is no need to do that. The distance function is a relation between pairs of points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ and a number r that is the distance between them. The points $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$, and the number r are related when

$$0 \leq r \wedge (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 = r^2.$$

The distance function for $\mathbb{R}^n$ can be identified with a definable relation on the set $\mathbb{R}^{2n+1}$. This makes all Euclidean geometry a part of the first-order structure of $\overline{\mathfrak{R}}$.

Both projections of the ellipse E in our example above are intervals. It turns out that if a subset X of $\mathbb{R}^n$ is defined by a first-order formula $\varphi(x_1, x_2, \ldots, x_n)$ of the language of $\overline{\mathfrak{R}}$, then the projection of X onto $\mathbb{R}$, defined by $\exists x_2 \cdots \exists x_n \varphi(x_1, x_2, \ldots, x_n)$, is a finite union of intervals. It is not difficult to

[4] The distance between two numbers a and b in $\mathbb{R}$ is defined to be $\sqrt{(a-b)^2}$, which can also be defined more naturally as $|a - b|$.

see that this is the case of Boolean combinations of conic sections. In general, it is a theorem proved independently by Abraham Seidenberg and Alfred Tarski. The crucial part of the proof is to show that if a subset X of R^n, with $n > 1$, is definable in $\overline{\mathfrak{R}}$ by a formula with no quantifiers, then its projection obtained by appending an existential quantifier in front of the formula, is again definable by a formula without quantifiers. It follows that in $\overline{\mathfrak{R}}$ every formula is equivalent to a formula without quantifiers. In the analysis of the geometry of sets that are definable in $\overline{\mathfrak{R}}$ one only has to consider sets definable by such formulas. The quantifiers have been eliminated. This is a actually a technical term: we say that $\overline{\mathfrak{R}}$ admits *elimination of quantifiers*.

Our statement of the Tarski-Seidenberg theorem is not quite correct. Everything is fine if we assume that the language of $\overline{\mathfrak{R}}$ includes the relation symbol $<$. Previously we have noted that one does not need to extend the language of $\overline{\mathfrak{R}}$ to include $<$. Addition and multiplication suffice because the ordering relation $<$ is definable in $\mathfrak{R}$. Now we have to be more careful. The defining formula is $a < b$ iff $\exists z[\neg(z = 0) \wedge (a + z \cdot z = b)]$. It has an existential quantifier. That quantifier cannot be eliminated. It can be shown that every definition of the ordering of $\mathbb{R}$ must use at least one quantifier. The Tarski-Seidenberg elimination of quantifiers for $\overline{\mathfrak{R}}$ uses the ordering in an essential way.

The discussion above focused on existential quantifiers. What about the universal ones? Universal quantifiers are indispensable in expressing various properties of structures in a natural way, but they can always be eliminated: instead of $\forall x \varphi(x)$ we can say: $\neg\exists x \neg\varphi(x)$. This shows that if all existential quantifiers can be eliminated in definitions of sets over a structure, then universal quantifiers can be eliminated as well.

10.3 Shadows and Complexity

Intuitively, the image under projection of a set from a higher dimension to lower should be less complex than the set itself. After all, all features of the shadow derive from the object, and some particular features may get lost. This is an example of how geometric intuition can fail us. Henri Lebesgue was one of the two most prominent mathematicians of the first quarter of the twentieth century.[5] The history of mathematics records a famous Lebesgue error concerning Borel sets.[6] Borel sets

[5] The other one was David Hilbert.

[6] If X is a metric, or in general a topological space, then the set of all Borel sets of X is the smallest set that contains all open subsets of X and is closed under complements, and countable unions and intersections i.e. if A is Borel, then so is $X \setminus A$, and if $A_1, A_2 \ldots$ is a sequence of Borel sets, then the union and the intersection of all sets in that sequence are also Borel.

can be quite complex, but from a certain point of view they are considered relatively simple. Lebesgue thought that the image under projection of a Borel subset of a plane $\mathbb{R}^2$ onto any of the coordinate axes is also Borel. So Lebesgue thought that the shadow of a relatively simple set is relatively simple, and he even gave a (false) proof of it. Lebesgue's error was found and corrected by Suslin, who showed that the image under projection of a relatively simple set in $\mathbb{R}^2$ can be much more complex than the projected set.

In the logic approach, we measure complexity of a definable set by the smallest number of quantifiers needed for its definition. Sets that can be defined without quantifiers have complexity 0, sets that can be defined by a formula with only one quantifier have complexity 1, and so on. In the case of the field of real numbers in the language including the relation symbol $<$, by the theorem of Tarski and Seidenberg all quantifiers can be eliminated, hence all definable sets have complexity 0. In a sharp contrast, in seemingly simpler structures $(\mathbb{N}, +\cdot)$, $(\mathbb{Z}, +\cdot)$, $(\mathbb{Q}, +\cdot)$, for each n there are sets that can be defined by formulas with $n + 1$ quantifiers, but cannot be defined by any formula with n quantifiers. A full explanation would require mathematical tools beyond the scope of this book, but in the following subsection we will discuss examples that may shed some light on where the complexity of those three structures comes from.

Let $\overline{\mathfrak{Z}}$ be the structure $(\mathbb{Z}, +, \cdot)$ expanded by adding names for all integers. All integers are definable $(\mathbb{Z}, +, \cdot)$, which shows that every relation definable in $\overline{\mathfrak{Z}}$ is already definable in $(\mathbb{Z}, +, \cdot)$, but since eliminating parameters in formulas by replacing them by their definitions adds additional quantifiers, and since we are paying close attention to quantifier complexity of formulas, now we will throw all the parameters into the language.

A bit surprisingly, $\overline{\mathfrak{Z}}$ turns out to be much more complex than $(\mathbb{R}, +, \cdot)$. In fact, $\overline{\mathfrak{Z}}$ is one of the most complex structures in mathematics.

To understand better the peculiar structure of $\overline{\mathfrak{Z}}$, and how is differs from $\overline{\mathfrak{R}}$ we need to discuss solvability of polynomial equations. We will do it in the next section by examining some prominent examples.

10.3.1 Diophantine Equations and Hilbert's 10th Problem

Let us begin with the equation $x^2 + y^2 = 1$. This equation has infinitely many real solutions. The solutions form a circle in $\mathbb{R}^2$ that is centered at $(0, 0)$ and has radius 1. Only four solutions have integer coordinates. They are (0,1), (1,0), (0,-1), and (-1,0). By analogy, one is tempted to say that those four points *are* a circle in $\mathbb{Z}^2$.

We will focus on the question whether a given polynomial equation has solutions in $\mathbb{Z}$, or, in other words, whether the set defined by the equation in $\overline{\mathfrak{Z}}$ is nonempty.

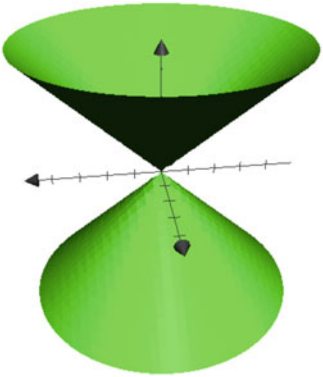


Fig. 10.1 The graph of $x^2 + y^2 = z^2$

The equation $x^2 + y^2 = 2$ also has four solutions in $\mathbb{Z}^2$: (1,1), $(-1, -1)$, $(-1, 1)$, and $(1, -1)$, but $x^2 + y^2 = 3$ has none, which can be quickly verified by direct checking, since the only integer candidates for x and y are 1, -1, and 0. The the last equation defines the circle in $\mathbb{R}^2$ that is centered at (0,0) and has radius $\sqrt{3}$, but each point (x, y) of that circle has at least one non-integer coordinate.

Let us now consider $x^2 + y^2 = z^2$. In $\mathfrak{N}$, this equation defines a subset of R^3. The solution set consists of two symmetric cones (Fig. 10.1). Some of the points in the solution set are in $\mathbb{Z}^3$, for example (0, 0, 0), or (3, 4, 5). In fact, the equation has infinitely many integer solutions. For every integer k, $(3k, 4k, 5k)$ is in the solution set, but there are many other.

Integer solutions of the equation $x^2 + y^2 = z^2$ are called *Pythagorean triples*. They are all known and classified, but this classification involves a good deal of elementary number theory, and it shows that the set defined by the equation in $\overline{\mathfrak{Z}}$ is quite complex.

The next example is $x^3 + y^3 = z^3$. This equation has trivial solutions, such as (0, 0, 0) or $(-2, 2, 0)$, but it has no solutions in which x, y, and z are positive integers (Fig. 10.2).

While there are many squares that are sums of two squares, the sum of two positive cubes can never be a cube.[7] In fact, for any n greater than 2, the equation $x^n + y^n = z^n$ has no positive integer solutions. This was observed by Pierre de Fermat (1601 or 1607–1665), the French lawyer and mathematician who on the margin of Diophantus' *Arithmetica* famously wrote "It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain." This statement, called Fermat's Last Theorem, was finally proved in 1994 by Andrew Wiles. The proof is immensely difficult, and it relies on a whole range of deep results from several areas of mathematics. For us, it offers a caveat: deciding whether sets defined

[7] A square is a number of the form n^2, and a cube is a number of the form n^3.

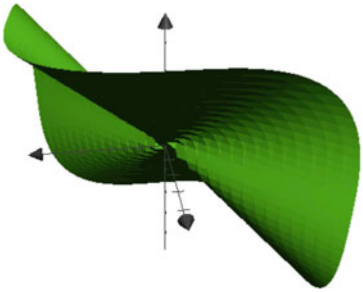


Fig. 10.2 The graph of $x^3 + y^3 = z^3$

by seemingly simple formulas, such as

$$x > 0 \wedge y > 0 \wedge z > 0 \wedge x^3 + y^3 = z^3,$$

are nonempty may require a lot of hard work.[8]

What makes such a big difference between $x^2 + y^2 = z^2$ and $x^n + y^n = z^n$, for $n > 2$? The difference is striking, but mathematical reasons for it are not easy to sort out. The syntax of the two formulas is almost identical, but the sets that the formulas define in $\mathfrak{Z}$ are very different. It is hard to know what the solution sets are just by just analyzing the form of their definitions.

There is a whole area of number theory, known as the theory of Diophantine equations,[9] that is devoted to the study of integer solutions of polynomial equations. Much of it consists of a painstaking analysis of particular cases. In this context, one should appreciate the audacity of David Hilbert who, addressing the International Congress of Mathematicians in Paris in 1900, included among the 23 challenge problems for the twentieth century mathematics his problem number 10: "Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers." In other words, Hilbert was asking if there is a mechanical procedure to decide whether a polynomial equation of the kind we are discussing has integer solutions. The meaning of the question at the time of Hilbert was equivocal, since the notion of a process that can be used to decide a mathematical question in a finite number of steps, while intuitively clear, had not been formalized yet.

Hilbert could not know that analogous question for the field of real numbers $\mathfrak{R}$ has a positive solution—the Tarski-Seidenberg theorem was proved only in the 1940s—but certain partial decision procedures of the kind Hilbert was asking for

[8] It was known well before Wiles' proof that the equation $x^3 + y^3 = z^3$ has no positive integer solutions. This was proved by another great mathematician Leonard Euler (1707–1783). Many other cases of Fermat's Last Theorem had been proved before Wiles announced his result. The smallest exponent n for which the theorem had not been verified before, is 4,000,037. This last fact is not very well-known. I am grateful to my colleague Cormac O'Sullivan for digging it up.

[9] After the third century Greek mathematician Diophantus of Alexandria.

had been known for a long time. To give a high school example, to find out if the equation $ax^2 + bx + c = 0$, where a, b, and c are integers, has a solution in $\mathbb{R}$, one can compute the number $\Delta = b^2 - 4ac$. The equation has a solution in $\mathbb{R}$ if and only if $\Delta \geq 0$, and has a solution in $\mathbb{Q}$ if and only if Δ is a square. Hilbert was asking for something much more general. The mechanical procedure he was asking for would give the answer for any polynomial equation with integer coefficients and any number of variables, and it would tell us if there are solutions in $\mathbb{Z}$.

Fermat's Last Theorem had to wait for its proof roughly 360 years. Hilbert's 10th problem was solved 70 years after it was posed, and the solution was negative. The solution is the celebrated MRDP theorem, proved in a collaborative effort over many years by Martin Davis, Yuri Matiyasevich, Hilary Putnam, and Julia Robinson.[10]

To see what the MRDP theorem has to do with shadows of definable sets, let us look again at equations defining conic sections in $\mathbb{R}^2$. We have to go through some technicalities, and we need some more notation.

Recall that a conic section in $\mathbb{R}^2$ is the set of points (x, y) defined in $\overline{\mathfrak{R}}$ by the formula

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \qquad (*)$$

where the parameters a, b, c, d, e, and f can be any real numbers. The solution set could be empty. For example, when $a = b = f = 1$, and all other parameters equal 0, the resulting equation is $x^2 + y^2 + 1 = 0$. This equation has no solutions, as the value on the left-hand side for any x and y is always at least 1.

Now the notation will become a bit more complicated. If this were a mathematics textbook, we would promptly introduce more notation so that the formulas we want to examine would not look that daunting. This is a good practice, but it takes time to learn the abbreviations, and to understand what they hide, so we will not do it here. The formulas will look rather complex, but what we are going to do is rather straightforward. We will make a small change in $(*)$ by replacing all parameters by variables. The equation $(*)$ becomes

$$z_1x^2 + z_2xy + z_3y^2 + z_4x + z_5y + z_6 = 0. \qquad (**)$$

Further, let $p(z_1, z_2, z_3, z_4, z_5, z_6, x, y)$ be the polynomial $z_1x^2 + z_2xy + z_2y^2 + z_4x + z_5y + z_6$. We are splitting hairs here, the only difference between $(*)$ and $(**)$ is that in the former a, b, c, d, e, and f represent arbitrary but fixed parameters in the equation; hence $(*)$ is a general form of an equation with two unknowns x and y. It represents infinitely many such equations, one for each selection of parameters. In $(**)$, we replaced the parameters by variables, so now it is a single equation with eight unknowns.

[10] For a very touching personal account of the history of the MRDP theorem see [29]. Technical details are included.

Let S be the subset of $\mathbb{R}^6$ defined by the formula

$$\exists x \exists y [p(z_1, z_2, z_3, z_4, z_5, z_6, x, y) = 0].$$

The quantifier-free formula

$$p(z_1, z_2, z_3, z_4, z_5, z_6, x, y) = 0$$

defines a subset A of $\mathbb{R}^8$. The set S is the image of A under the projection onto $\mathbb{R}^6$. According to the Tarski-Seidenberg theorem, S has a quantifier-free definition, which means that there is a polynomial $q(z_1, z_2, z_3, z_4, z_5, z_6)$ such that S is the set of those $(z_1, z_2, z_3, z_4, z_5, z_6)$ in $\mathbb{R}^6$ for which $q(z_1, z_2, z_3, z_4, z_5, z_6) = 0$.

The Tarski-Seidenberg theorem also provides an algorithmic procedure that given the polynomial p as above, produces q with the required property. Now, for the given integer parameters a, b, c, d, e, and f, one can compute $q(a, b, c, d, e, f)$ and if the result is 0, this means that the equation $(*)$ with those numbers as parameters has a solution in $\mathbb{R}^2$, and otherwise it does not.

Compare the two definitions of the set S. The first one employs a definition with two existential quantifiers; the second is quantifier-free. Deciding if the equation $(*)$ has a solution using the first definition involves a potentially infinite search for x and y; hence, while formally well-defined, it may be of little use. Searching for suitable x and y in a vast infinite domain is, in general, not a feasible task. In contrast, the second definition involving the polynomial q is based on a straightforward computation.[11]

We have outlined a decision procedure for solvability of arbitrary quadratic equations in two unknowns x and y. The same procedure applies to all polynomial equations in any number of variables.

Let us now see what changes when instead of solvability of equations in $\mathbb{R}$ one is interested in solutions in $\mathbb{Z}$. One of the consequences of the MRDP theorem is that there is a polynomial

$$h(z_1, z_2, \ldots, z_m, x_1, x_2, \ldots, x_n),$$

in which the variables $z_1, z_2, \ldots, z_m$ represent parameters, and variables $x_1, x_2, \ldots, x_n$ represent unknowns, such that there is no algorithmic procedure to decide whether for given integer parameters $a_1, a_2, \ldots, a_m$, there are integers $k_1, k_2, \ldots, k_n$ such that

$$h(a_1, a_2, \ldots, a_m, k_1, k_2, \ldots, k_n) = 0. \tag{†}$$

[11] In practice it may not be as simple as it appears, and in fact quite often it is not. The polynomial $q(z_1, z_2, z_3, z_4, z_5, z_6)$ may be of high degree, and calculations needed to evaluate it may be tedious, or simply too hard, even for a fast computer.

Contrary to what Hilbert himself believed, his 10th problem not only has a negative solution in general; the solution is negative even in the case of parametric versions of a single polynomial.

Much advanced mathematics goes into details of the results in this section. In particular, the construction of the polynomial h is messy, and there is still research going on to minimize both the number of parameters in $z_1, z_2, \dots, z_m$, and the number of variables in $x_1, x_2, \dots, x_n$ in it.

Now we can go back to projections. Let S be the set of those parameters $a_1, a_2, \dots, a_m$, for which the equation (†) above has a solution $(k_1, k_2, \dots, k_n)$ in $\mathbb{Z}^n$. Then, S is a subset of $\mathbb{Z}^{m+n}$, and it is defined in $\overline{\mathfrak{Z}}$ by a formula with n existential quantifiers

$$\exists x_1 \exists x_2 \cdots \exists x_n [p(z_1, z_2, \dots, z_m, x_1, x_2, \dots, x_n) = 0].$$

We cannot replace this definition with an equivalent quantifier-free formula, because if we could, that would give us a computational procedure to check membership in the set S, but we know from the MRDP theorem that there is no such procedure. So this is our example. The set A of all $(z_1, z_2, \dots, z_m, x_1, x_2, \dots, x_n)$ such that

$$p(z_1, z_2, \dots, z_m, x_1, x_2, \dots, x_n) = 0$$

is simple, because a calculation involving only addition and multiplication of integers can reveal if a point in $\mathbb{Z}^{m+n}$ is in it or not. Its projection S is much more complex. A given sequence $(a_1, a_2, \dots, a_m)$ may be in A or may be not, but we do not have any way of checking by a computation.

So here is our conclusion: in $\overline{\mathfrak{Z}}$ the image under projections of a simply defined set can be much more complex then the set itself. The interesting part of this story is that this syntactic complexity of the set, as measured by the number of quantifiers in its definition does not determine the complexity of projected images. The final outcome depends on semantics; it depends on the structure in which the language is interpreted. First-order theories of $\overline{\mathfrak{Z}}$ and $\overline{\mathfrak{R}}$ share the same syntax, but their semantics are very different.

10.3.2 The Reals vs. The Rationals

One could object that the comparison between $\overline{\mathfrak{Z}}$ and $\overline{\mathfrak{R}}$ is unfair. Why should one expect the geometries of definable sets in those two structures to be similar? Just by looking at the ordered sets $(\mathbb{Z}, <)$ and $(\mathbb{R}, <)$ one sees stark differences. The former is a countable (small) discrete order, while the latter is uncountable (huge) and dense. With a few exceptions, each polynomial equation in two unknowns defines a smooth curve in $\mathbb{R}^2$, but only rarely do those curves pass through points both of whose coordinates are integers. It seems reasonable to expect that the geometry of $\mathbb{Z}^2$ is much more complex than that of $\mathbb{R}^2$, as indeed it is.

Algorithms for addition and multiplication of integers represented in decimal notation are straightforward, but the simplicity of these operations is deceptive. In polynomial equations, addition and multiplication are mixed together, and this creates complexity. As we have seen, this complexity disappears when we move to the much smoother structure $\overline{\mathfrak{R}}$. The transition from $\mathbb{Z}$ to $\mathbb{R}$ proceeds via the intermediate structure of rational numbers $\mathbb{Q}$, so now we need to discuss how complex are the definable sets in that structure.

Many equations with integer coefficients, such as $x + x = 1$, do not have integer solutions. The set defined in $\overline{\mathfrak{Z}}$ by $x + x = 1$ is empty, but in $(\mathbb{Q}, +)$ it defines the one element set $\{\frac{1}{2}\}$. Many more equations can be solved in $\mathbb{Q}$ than in $\mathbb{Z}$, but not all. For example $x \cdot x = 2$, has no solutions in $\mathbb{Q}$, but it has two solutions in $\mathbb{R}$. The set of rational numbers defined by this equation is empty, but in $\overline{\mathfrak{R}}$ it defines the two element set $\{-\sqrt{2}, \sqrt{2}\}$. In the evolution of the number system we moved from the simpler and easier to describe set $\mathbb{Q}$ to a much larger and somewhat mysterious $\mathbb{R}$ exactly for this reason—to be able to find all solutions to polynomial equations that represent geometric quantities. The domain $\mathbb{R}$ is incomparably more complex than the simple set $\mathbb{Z}$, but the additive and multiplicative structure of the real numbers is much smoother, and in effect the geometry of parametrically definable sets becomes much easier to describe.

The domain $\mathbb{Q}$ is countable, so not as large as $\mathbb{R}$, but the ordering is dense, and it is a field. Let $\overline{\mathfrak{Q}}$ be $(\mathbb{Q}, +, \cdot)$ in the language with names for all rational numbers. On the one hand, because of the density of its order, $\overline{\mathfrak{Q}}$ looks a bit like $\overline{\mathfrak{R}}$, on the other hand, because many polynomial equations do not have rational solutions, it inherits some complexity from $\overline{\mathfrak{Z}}$. In terms of logical complexity $\overline{\mathfrak{Q}}$ is somewhere between $\overline{\mathfrak{Z}}$ and $\overline{\mathfrak{R}}$, but it turns out that $\overline{\mathfrak{Q}}$ is nowhere near $\overline{\mathfrak{R}}$.

The geometry of definable relations on $\overline{\mathfrak{Q}}$ is as complex as that of $\overline{\mathfrak{Z}}$, but this is neither clear nor easy to prove. It follows from a theorem of Julia Robinson, proved in 1949. Robinson found a formula that defines the set $\mathbb{Z}$ in $\overline{\mathfrak{Q}}$. This showed that to be an integer is a first-order property in the field of rational numbers. The great complexity of $\overline{\mathfrak{Z}}$ is present in $\overline{\mathfrak{Q}}$. This statement may seem obvious, since $\mathbb{Z}$ is a subset of $\mathbb{Q}$, but the fact that $\mathbb{Z}$ is logically visible $\overline{\mathfrak{Q}}$ makes $\mathbb{Z}$ a part of the geometry of the field of rational numbers. Recall that, by the Tarski-Seidenberg theorem, $\mathbb{Z}$ is not parametrically definable in the field of real numbers.

The field of rational numbers is at least as complex as $\overline{\mathfrak{Z}}$, but it is not more complex than $\overline{\mathfrak{Z}}$. When we introduced the structure $(\mathbb{Q}, +, \cdot)$, we did it by defining it in a first-order way in $(\mathbb{Z}, +, \cdot)$, hence the former structure is logically no more complex than the latter: any relation that is logically visible in $(\mathbb{Q}, +, \cdot)$ is also visible in $(\mathbb{Z}, +, \cdot)$, and vice versa. From the point of view of logical analysis, $\overline{\mathfrak{Z}}$ and $\overline{\mathfrak{Q}}$ are equivalent.

Exercises

Exercise 10.1 *Show that every integer is definable in the structure* $(\mathbb{Z}, +, \cdot)$*. Write explicit definitions of* 3 *and* -3*, try to use as few quantifiers as possible. What is the quantifier complexity of your definition? Hint: See Exercise* 9.2.

Exercise 10.2 *By direct checking, find all integer solutions of the equation* $x^2 + y^2 = 4$*, and show that* $x^2 + y^2 = 3$ *has none.*

Exercise 10.3 * *Express the Pythagorean theorem by first-order sentence in the language with* $+$ *and* $\cdot$*. Hint: Let* $A = (a_1, a_2)$*,* $B = (b_1, b_2)$*, and* $C = (c_1, c_2)$ *be points in* $\mathbb{R}^2$*. Then the segments* BA *and* BC *form a right angle if and only if*

$$(a_1 - b_1) \cdot (c_1 - b_1) + (a_2 - b_2) \cdot (c_2 - b_2) = 0.$$

Exercise 10.4 * *Suppose that for every quantifier-free formula* $\varphi(x, x_1, \dots, x_n)$ *of the language of a structure* $\mathfrak{A}$*, there is a quantifier-free formula* $\psi(x_1, \dots, x_n)$ *such that for all* $a_1, \dots a_n$ *in the domain of* $\mathfrak{A}$*,* $\exists x \varphi(x, a_1, \dots, a_n)$ *holds in* $\mathfrak{A}$ *if and only if* $\psi(a_1, \dots, a_n)$ *holds in* $\mathfrak{A}$*. Show that every sentence* φ *of the language of* $\mathfrak{A}$ *is equivalent to a quantifier-free sentence of that language. Hint: Use induction on the logical complexity of* φ*.*