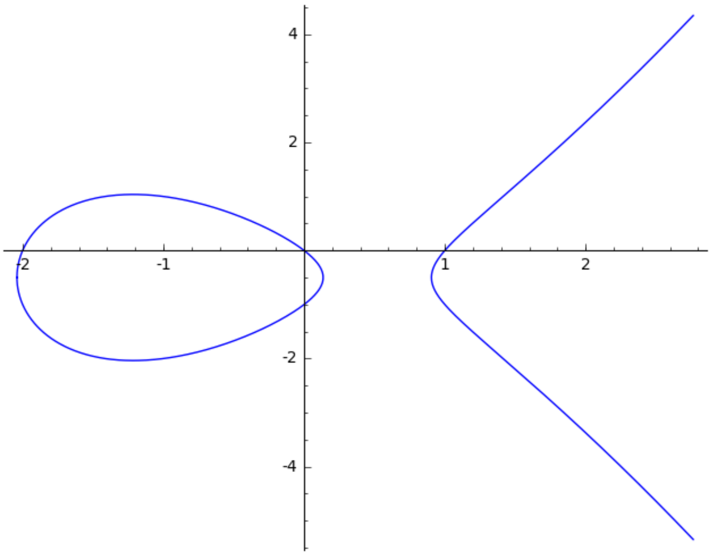
Doç. Dr. Hatice Boylan

Number Theory II



Lecture Notes
İstanbul Üniversitesi

Version: 10.03.16

Contents

Foreword

The following speech, given 30 years ago at the occasion of the opening of one of the most famous research institutes of the world for mathematical sciences might help the reader to come closer to answer the questions *'What is number theory?'* and *'What is it good for?'*:

> ## Why do you study number theory?
>
> Mathematics and German share the same disadvantage, both are universally applicable and at the same time they are the summit of artistic creation of human kind. Why do we need Goethe if we can express our wishes clearly at the market place? And for what do we need number theory if we can solve the differential equation of the heat equation numerically? Strangely enough, in this competition those domains do better which have no imaginable commercial application. One of my colleagues at Durham University was once asked by the local TV why he studies the precise dating of Crete vases, and he answers that this would be very useful for the study of the migration of the Minoan civilization. To my surprise this was accepted with respectful appreciative murmuring.

Hence our first answer to the question 'Why do you study number theory' should possibly be 'It is indispensable for the right understanding of modular forms.' After we have now put down the objections of the trifling and superficial people, we can try to answer seriously. The serious answer is, of course: 'Why not?'. Namely, beavers build dams and cuckoo borrow nests without any intent of refunding, but only humans (as far as we know) worry about the questions which prime numbers are the sum of two squares. Since we reached a partial freedom from the urgent need of surviving, the desire for knowledge and the expression of beauty were always the ultimate goal of the human race. The purpose of technology and invention is to give us time for the further study of Bach, Gauß and Goethe, and not vice versa. But it is one of the divine compensations for our existence that the compulsive quest for knowledge almost always eventually carries practical fruits.

A.O.L. Atkin. Chicago. at the occasion of a visit to Max-Planck-Institut for Mathematics. Bonn. June 1985.

This is a free translation of the following German text, which appeared in one of the internal publications of the Max-Planck Gesellschaft and was in turn probably translated from the original English speech. We are grateful for any hint to the original speech.

Warum studieren Sie Zahlentheorie?

„Mathematik hat mit Deutsch den Nachteil gemeinsam, sowohl universell anwendbar wie zugleich einer der Gipfel des künstlerischen Schaffens der Menschheit zu sein. Wozu braucht man Goethe, wenn man am Marktplatz seine Wünsche klar ausdrücken kann? Und wozu braucht man die Zahlentheorie, wenn man die Differentialgleichungen der Wärmeleitung numerisch lösen kann? Merkwürdig genug fahren bei diesem Spiel die Gebiete, die keine denkbare kommerzielle Anwendung haben, oft viel besser. Einer meiner Kollegen an der Durham University wurde einmal vom lokalen Fernsehen gefragt, warum er die Theorie der präzisen Datierung kretischer Vasen studiere, und antwortete, dies sei sehr nützlich beim Studium der Migration der minoischen Zivilisation. Zu meiner Überraschung wurde dies mit respektvollem Anerkennungsgemurmel akzeptiert.

Also soll unsere erste Antwort auf die Frage „Warum studieren Sie Zahlentheorie?“ vielleicht „Sie ist für das richtige Verständnis der Modulformen unentbehrlich“ sein. Nachdem wir nun die Einwände der Frivolen und Oberflächlichen erledigt haben, können wir versuchen, ernsthaft zu antworten. Die ernsthafte Antwort ist natürlich: „Warum nicht?“ Denn Biber bauen Dämme, und Kuckucks leihen Nester ohne jegliche Rückzahlungsabsicht, aber nur Menschen (soweit wir wissen) zerbrechen sich den Kopf über die Frage, welche Primzahlen Summen zweier Quadratzahlen sind. Das Verlangen nach Wissen und der Ausdruck von Schönheit sind, seitdem eine partielle Freiheit von dem niederen Überlebensbedürfnis erreicht wurde, immer das höchste Ziel der menschlichen Rasse gewesen. Der Zweck von Technologie und Erfindung ist es, unsere Zeit für das weitere Studium von Bach, Gauß und Goethe freizumachen, und nicht umgekehrt. Es ist aber eine der göttlichen Entschädigungen für unser Dasein, daß die zwanghafte Suche nach Wisssen fast immer später praktische Früchte trägt.“

A. O. L. Atkin, Chicago, anläßlich eines Besuchs im MPI für Mathematik, Juni 1985.

Preface

These lecture notes represent the optional course *Number Theory II*, a continuation of the obligatory course *Number Theory I.* The lecture notes for the Number Theory I consisted of 4 sections (available at *Ders Notu Satış Bürosu*). Accordingly the present lecture notes start with Sections 5. Though the current study regulations do not ask for giving the second course in English we preferred to continue these lecture notes in English since we felt that it would be an additional burden for the students to switch from the English terminology as learned in the first part to the Turkish one.

Hatice Boylan İstanbul Üniversitesi, March 2016

Chapter 2

Higher Methods

5. Arithmetical functions

There are several functions $f(n)$ depending on a non-negative integer n which occur naturally in number theory. Examples are the number $\varphi(n)$ of primitive residue classes modulo n, the sum $d(n)$ of the divisors of a number n, the number of primes dividing n (say, including multiplicities) and the like. Many of these functions share properties which are useful in various situations and which we shall study in this section. Though an *arithmetical function* is usually a map $f : n \mapsto f(n)$ which is somehow motivated by arithmetical considerations it is useful to simply adopt the following definition.

Definition. An *arithmetic function* is a map $f : \mathbb{Z}_{\geq 1} \to \mathbb{C}$.

Example. Examples of arithmetic functions are:

(1) The *Euler φ-function* which associates to n the number $\varphi(n)$ of primitive residue classes modulo n,

(2) the *divisor sum* which associates to n the sum $d(n)$ of the divisors of n,

(3) and, more generally, for any given k, the function σ_k whose values $\sigma_k(n)$ equal the sum of the kth powers of the divisors of n,

(4) the Liouville-function λ, where $\lambda(n) = (-1)^{\Omega(n)}$ and $\Omega(n)$ equals the number of prime factors of n, counted with multiplicities.

Several of these examples are obtained by summing a given simple function over the divisors of n. More formally, we call G the *summatory function of g* if

$$G(n) = \sum_{d|n} g(d).$$

In such an expression we always mean that d runs over the positive divisors of n. In Table 1 the function in the row G is always the summatory function of the one below in the row g.

$G(n)$	$d(n)$	$\sigma(n)$	$\sigma_k(n)$	n
$g(n)$	1	n	n^k	$\varphi(n)$

Table 1. Summatory functions $G(n) = \sum_{d|n} g(d)$

Another observation is that the functions of all the above examples are *multiplicative*, which means that a given function f is not identically 0 and

$$f(mn) = f(m) \cdot f(n) \quad \text{for all } m,\, n \text{ such that } \gcd(m,n) = 1.$$

We proved this already for Euler's φ-function and it is not hard to verify this for the other examples. However we shall prove in a moment a theorem which makes it easy to recognize this property. An f which satisfies $f(mn) = f(m) \cdot f(n)$ for all m and n without any restriction is called *strongly multiplicative.* The Liouville λ-function is obviously strongly multiplicative. Note that a multiplicative function always satisfies $f(1) = 1$: indeed $f(1) = f(1)\,f(1)$, and $f(n) = f(1)\,f(n)$; the latter implies that $f(1) \neq 0$ (since f must not be identically 0), and the former then our claim. Also, for a multiplicative function one has

$$f(n) = \prod_{p^\nu \| n} f(p^\nu).$$

A already used before this writing means that the product is to be taken over all maximal prime powers dividing n. Applying this to the function σ_k gives for example

$$\sigma_k(n) = \prod_{p^\nu \| n} (1 + p^k + \cdots + p^{k\nu}) = \prod_{p^\nu \| n} \frac{p^{k(\nu+1)} - 1}{p^k - 1}.$$

5.1. The ring of arithmetic functions. Many of the properties of arithmetic functions are best understood in terms of a natural structure of a ring which one can define on them. As usual we can add two arithmetic functions f and g by defining their sum $f+g$ by $(f+g)(n) = f(n) + g(n)$.

Definition. The *Dirichlet product of to arithmetic functions f and g* is the arithmetic function $f * g$ which is defined by

$$(f * g)\,(n) = \sum_{d|n} f(d)\,g(n/d) = \sum_{de=n} f(d)\,g(e).$$

(The second sum is over all pairs (d, e) of positive integers such that $de = n$.)

As a first indicator for the usefulness of these definitions note that, for a given arithmetic function f the summatory function F is nothing else but $f * C$, where C is the functions with single value 1 (i.e. $C(n) = 1$ for all n).

Theorem. *The set $\mathfrak{A}$ of arithmetic functions together with the usual addition and the Dirichlet product satisfies the axioms of a commutative ring with neutral element.*

Proof. We have to show that for our addition and Dirichlet product the following properties are satisfied.

(1) $f + (g + h) = (f + g) + h$

(2) $f + g = g + f$

(3) $f + 0 = f$

(4) $f + (-f) = 0$

(5) $f * (g * h) = (f * g) * h$

(6) $f * g = g * f$

(7) $f * \mathbb{E} = f$

(8) $f * (g + h) = f * g + f * h$

Here 0 denotes the function which is identically 0, and $-f$ is the function such that $(-f)(n) = -f(n)$. Finally, $\mathbb{E}$ denotes the function

$$\mathbb{E}(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{otherwise.} \end{cases}.$$

We leave it to the reader to verify these properties. Most of them are quickly checked. For (5) we suggest to verify that both sides, evaluated at an argument n, equal

$$\sum_{abc=n} f(a)g(b)h(c),$$

the sum being over all triples (a, b, c) of positive integers such that $abc = n$. □

Though it is not necessary for the understanding of the following it might be helpful for the interested reader to look at the subsequent results with a bit of abstract algebra. Given a ring R one is interested in the group R^* of units in R. By this one means the subset R^* of elements r in R for which there exists an element s in R such that $rs = sr = 1$ (where 1 denotes the multiplicative unit element in R). If r and r' are units then rr' is a unit, and, in fact, R^* possesses the structure of a group with respect to the multiplication in the ring.

Theorem. *Let $f \in \mathfrak{A}$. Then $f \in \mathfrak{A}^*$ (i.e. there exists a g in $\mathfrak{A}$ such that $f * g = \mathbb{E}$) if and only if $f(1) \neq 0$.*

Proof. If $f * g = \mathbb{E}$ for some g, then in particular $f(1)g(1) = 1$, and therefore $f(1) \neq 0$. Assume vice versa that $f(1) \neq 0$. We define g by induction: set $g(1) = 1/f(1)$ and

$$g(n) = -\frac{1}{f(1)} \sum_{\substack{d|n \\ d<n}} g(d)\, f(n/d).$$

Clearly then $(g * f)(1) = 1$ and $(g * f)(n) = 0$ for $n \geq 2$. □

Theorem. *The set $\mathfrak{M}$ of multiplicative arithmetic functions is a subgroup of $\mathfrak{A}^*$, i.e.:*

(i) *if f is multiplicative, then $f \in \mathfrak{A}^*$,*

(ii) *if f and g are multiplicative, then $f * g$ is multiplicative too,*

(iii) *if f is multiplicative, then its inverse f^{-1} (with respect to the Dirichlet product) is multiplicative.*

Proof. We saw already that a multiplicative function f satisfies $f(1) = 1$, so that, by the preceding theorem, it is invertible.

Assume that f and g are multiplicative and let $h = f * g$. For proving that h is multiplicative let m and n be two positive and relatively prime integers. We leave it to the reader to verify that the application $(d, e) \mapsto de$ defines a bijection

$$\mathfrak{D}(m) \times \mathfrak{D}(n) \xrightarrow{\cong} \mathfrak{D}(mn),$$

where, for an integer l, we use $\mathfrak{D}(l)$ for the set of divisors of l. Using this bijection we find

$$\begin{aligned} h(mn) &= \sum_{t \in \mathfrak{D}(mn)} f(t)\, g(mn/t) = \sum_{(d,e) \in \mathfrak{D}(m) \times \mathfrak{D}(n)} f(de)\, g(mn/de) \\ &= \sum_{(d,e) \in \mathfrak{D}(m) \times \mathfrak{D}(n)} f(d) f(e)\, g(m/d) g(n/e) \\ &= \sum_{d \in \mathfrak{D}(m)} f(d)\, g(m/d) \sum_{e \in \mathfrak{D}(n)} f(e)\, g(n/e) = h(m)h(n). \end{aligned}$$

The proof that f^{-1} is multiplicative if f is multiplicative is similar (use the formula for f^{-1} from the preceding proof). □

Example. From the theorem it is immediate that the *divisor sum functions* $\sigma_k = C * \mathrm{Id}^k$ are multiplicative, since the constant function $C \equiv 1$, the identity function Id and then also $\mathrm{Id}^k : n \mapsto n^k$ are obviously multiplicative.

If G is the summatory function of a given arithmetic function g it is often useful to be able to express g in terms of G. This is indeed always possible. Namely, that G is the summatory function of g means that $G = C * g$ with C denoting the constant function $n \mapsto 1$. Since $C(1) = 1$ we know that C is invertible. Hence $g = C^{-1} * G$. For turning this into a useful formula we need to study C^{-1}, which we shall do now.

Definition (Möbius' μ-function)**.** The arithmetic function μ which is defined by

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ is not squarefree,} \\ (-1)^r & \text{if } n \text{ is squarefree and the product of } r \text{ primes.} \end{cases}$$

is called the *Möbius μ-function.*

A number n is called *squareferee* if n is not divisible by the square of a prime, or, equivalently, by a perfect square different from 1.

n	1	2	3	4	5	6	7	8	9	10
$\mu(n)$	1	-1	-1	0	-1	1	-1	0	0	1

Table 2. The first values of the Möbius μ-function

Theorem. *For any positive integer n, one has*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & n = 1, \\ 0 & n > 1 \end{cases}$$

(i.e.. μ is the inverse of the constant function $C \equiv 1$ with respect to the Dirichlet multiplication).

Proof. Set $G(n) = \sum_{d|n} \mu(d)$, i.e. $G = C * \mu$. Clearly $G(1) = 1$. From the definition of μ it is clear that μ is multiplicative, and so is then G too. Hence, for proving $G(n) = 0$ for $n \geq 2$ it suffices to calculate $G(p^r)$ for any given prime power p^r. But from the definition of μ we have $G(p^r) = \mu(1) + \mu(p) = 0$,which proves the theorem. □

We can finally make the above formula $g = C^{-1} * G$ more explicit.

Theorem (Möbius' inversion)**.** *Let G and g be arithmetic functions. Then one has:*

$$\forall n : G(n) = \sum_{d|n} g(d) \quad \textit{if and only if} \quad \forall n : g(n) = \sum_{d|n} \mu(n/d)\, G(d).$$

Proof. This is an immediate consequence of the preceding theorem which can be restated as $\mu * C = \mathbb{E}$, i.e. $\mu = C^{-1}$. Hence $G = g * C$ if and only if $g = G * \mu$, which is the claim. □

Example. 1. We saw in Section **??** that $\sum_{d|n} \varphi(d) = n$ for all n. Via M"obius inversion we obtain the formula

$$\varphi(n) = \sum_{d|n} \mu(n/d)\, d.$$

Another type of inversion which one encounters often in number theory is the following. Let f be an arithmetic function and let, for any positive integer n,

$$F(n) = \sum_{\substack{d|n \\ n/d\text{squarefree}}} f(d).$$

Here the sum is over all positive divisors d of n such that n/d is squarefree. The above equation can be written shorter as $F = f * \chi_{\mathrm{sf}}$, where $\chi_{\mathrm{sf}}(n) = 1$ if n is squarefree and equals 0 otherwise. Since $\chi_{\mathrm{sf}}(1) = 1$ the function χ_{sf} is a unit, and hence $f = \chi_{\mathrm{sf}}^{-1} * F$. We leave it to the reader to prove

Theorem. *one has*

$$\chi_{\mathrm{sf}}^{-1} = \lambda.$$

(Here λ is Liouville's λ-function.)

In other words, the above formula expressing F in terms of f is equivalent to

$$f(n) = \sum_{d|n} F(d)\, \lambda(n/d).$$

5.2. Reinterpretation of the Dirichlet product. It is useful to go even a bit further in the application of tools from abstract algebra to the study of arithmetic functions. The reader inexperienced in algebra might want to skip this section. If f is an arithmetic function and z a complex number then zf, i.e. the function $n \mapsto zf(n)$, is also an arithmetic function. This multiplication of arithmetic functions by scalars (together with the addition of arithmetic functions) turns $\mathfrak{A}$ into a vector space over the complex numbers $\mathbb{C}$. In fact, $\mathfrak{A}$ equipped with this scalar multiplication, the addition of functions and the Dirichlet product satisfies the axioms of a *commutative algebra over* $\mathbb{C}$.

Let $\{f_i\}_{i\in I}$ be a (possibly infinite) family[1] of arithmetic functions such that for each integer $n \geq 1$ one has $f_i(n) = 0$ for all but finitely many i in I. We call such a family *summable.* We can then define the *sum of the family* $\{f_i\}_{i\in I}$, denoted by

$$\sum_{i\in I} f_i$$

as the arithmetic function which associates to a given integer $n \geq 1$ the (finite) sum $\sum_{i\in I} f_i(n)$.

Special summable families are obtained as follows. For an integer $n \geq 1$, we use n^{-s} for the arithmetic function which maps n to 1 and $n' \neq n$ to 0. The family $\{n^{-s}\}_{n\in\mathbb{Z}_{\geq 1}}$ is obviously summable, and so is $\{f(n)\, n^{-s}\}_{n\in\mathbb{Z}_{\geq 1}}$ for any given arithmetic function f. Using these notations we can now write every arithmetic function f as sum of the family $\{f(n)\, n^{-s}\}_{n\in\mathbb{Z}_{\geq 1}}$, i.e. we can write any f in the form

$$f = \sum_{n\geq 1} f(n)\, n^{-s}.$$

The expression on the right is called *a formal Dirichlet series.* The reader should note that we did not introduce a new mathematical object, but merely a new language for treating arithmetic functions. We shall see that this language has certain advantages.

We encourage the reader to practice calculating with formal Dirichlet series and thereby discover natural rules to manipulate them. The first thing to discover is that the Dirichlet product becomes very natural in the language of formal Dirichlet series. Namely, one has $m^{-s} * n^{-s} = (nm)^{-s}$. Because of this one usually uses the dot "$\cdot$" for the operation symbol $*$ when dealing with formal Dirichlet series, and sometimes one simply omits the dot, so that we can write, for example, $m^{-s}n^{-s} = (mn)^{-s}$. The product of two arithmetic functions

[1] A *family* $\{x_i\}_{i\in I}$ *of elements of a set* X is nothings else than a map $I \to X$, $i \mapsto x_i$.

f and g can then be calculated as

$$\begin{aligned}&\left(\sum_{m\ge1} f(m)\,m^{-s}\right)\left(\sum_{n\ge1} g(n)\,n^{-s}\right)\\ &\quad= \sum_{m,n\ge1} f(m)g(n)\,m^{-s}n^{-s} = \sum_{l\ge1}\left(\sum_{mn=l} f(m)g(n)\right) l^{-s}\end{aligned}$$

Here we use for the first identity that the product of the sums of two summable families $\{f_i\}_{i\in I}$ and $\{g_j\}_{j\in J}$ equals the sum of the (again summable) family $\{f_i * g_j\}_{(i,j)\in I\times J}$. For the second identity we use that the sum of the family $\{f_i\}_{i\in I}$ equals the sum of $\{\sum_{i\in I_j} f_i\}_{j\in J}$ for any partition of I into a disjoint union of finite sets I_j $(j\in J)$.

We call a (possibly infinite) family $\{f_i\}_{i\in I}$ of arithmetic functions *multiplicable* if we have

(1) for each $n\ge 2$, we have $f_i(n)=0$ for all but finitely many i,

(2) and $f_i(1)=1$ for all i.

For a multiplicable family we define the *product* $\prod_{i\in I} f_i$ *of the family* $\{f_i\}_{i\in I}$ as the arithmetic function

$$n\mapsto \sum_{\substack{\{d_i\}_{i\in I}\\ n=\prod_{i\in I} d_i}} \prod_{i\in I} f_i(d_i).$$

Here $\{d_i\}_{i\in I}$ runs through all families of positive integers d_i such that $d_i=1$ for all but finitely many i. The inner product has to be understood as the finite product over all i in I such that $f_i(d_i)\ne 1$. The sum is over the (finitely) many families $\{d_i\}_{i\in I}$ for which the inner sum is different from zero. The reader should notice that this product applied to a finite family is nothing else but the Dirichlet product of the members of this family.

Special multiplicable families are obtained as follows: for each prime p let a_p be a map $\mathbb{Z}_{\ge0}\to\mathbb{C}$ with $a_p(0)=1$. Then the family

$$\{\sum_{k\ge0} a_p(k)\,p^{-ks}\}_p$$

(where p runs through all primes) is multiplicable. Indeed, for any given $n\ge2$, $\left(\sum_{k\ge0} a_p(k)\,p^{-ks}\right)(n)\ne0$ for at most one p (namely, at most if n is power of p). The product of such a family is called an *Euler product*. The reader should verify that the product is the arithmetic function

$$n \mapsto \prod_{p^k \,\|\, n} a_p(k),$$

where the product is over all prime powers p^k exactly dividing n (i.e. dividing n such that n/p^k is relatively prime to p). The reader should also prove that an arithmetic function f is multiplicative if and only if it can be factored into an Euler product, which means that $f(1) = 1$ and

$$f = \prod_p \sum_{k \geq 0} f(p^k)\, p^{-ks}.$$

It is quickly verified that, for each prime p, the formal Dirichlet series $\sum_{k\geq 0} p^{-ks}$ is the inverse (with respect to Dirichlet multiplication) of $1 - p^{-s}$. Therefore

$$\begin{aligned}
C &= \sum_{n\geq 1} n^{-s} = \prod_p \frac{1}{1-p^{-s}}, \\
\mu &= C^{-1} = \prod_p \left(1 - p^{-s}\right), \\
\sigma_k &= \sum_{n\geq 1} \sum_{d|n} d^k\, n^{-s} = \prod_p \frac{1}{(1-p^{-s})(1-p^{k-s})}, \\
\varphi &= \prod_p \frac{1-p^{-s}}{1-p^{1-s}}, \\
\lambda &= \prod_p \frac{1}{1+p^{-s}}.
\end{aligned}$$

Note that from these formulas various identities which we encountered before become rather trivial. For example, the identity $\sum_{d|n} \varphi(d) = n$ reads in terms of formal Dirichlet series $C\varphi = \mathrm{Id}$, which is obvious from the Euler product decomposition of C and φ. Similarly, the identities $\mu = 1/C$ and $\lambda = 1/\chi_{\mathrm{sf}}$ become trivial when rewritten in terms of Euler products.

As an exercise the reader might try to find out how to describe the property of an arithmetic function f to be strongly multiplicative in terms of its Euler product.

Instead of dealing with formal Dirichlet series one can also use *real* Dirichlet series.

Definition (Dirichlet series)**.** A *Dirichlet series* is an infinite series of functions of the form

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s}$$

where f is an arithmetical function.

Recall that, for any n and s, one has $n^s = \exp(-s \log n)$. It arises now the natural question if there are, say real, numbers s, for which a given Dirichlet series converges. It is quickly satisfied that a given Dirichlet series converges absolutely for all $s \geq s_0$ if it converges absolutely for s_0, and it defines then a smooth function $D_f(s)$ in the interval $s > s_0$[2]. Examples for such Dirichlet series are provided by those f which do not grow faster than some polynomial, i.e. by those f for which there exist a constant c and an integer k such that $|f(n)| \leq cn^k$ for all $n \geq 1$. Here the infinite series $\sum f(n)/n^s$ is absolutely convergent for all $s > k + 1$.

The connection between the theory of Dirichlet series and of formal Dirichlet series is as follows. Let $\mathfrak{A}_{\text{ac.}}$ the set of arithmetic functions f such that $\sum_{n\geq 1} f(n)/n^s$ is absolutely convergent for some s_0. It is not hard to prove that $\mathfrak{A}_{\text{ac.}}$ is in fact a sub-algebra of $\mathfrak{A}$, and that the map $f \mapsto D_f$ defines an injective homomorphism of algebras from $\mathfrak{A}_{\text{ac.}}$ into the algebra of smooth functions which are each defined in some ray $s > s_0$ (depending on the function). The property of being a homomorphism of algebras means $D_{f+g} = D_f + D_g$, $D_{zf} = zD_f$, $D_{f*g} = D_f \cdot D_g$ for all f and g in $\mathfrak{A}_{\text{ac.}}$ and all complex numbers z. Note that $D_f \cdot D_g$ denotes here the usual product of functions (i.e. $(D_f \cdot D_g)(s) = D_f(s)D_g(s)$).

The Dirichlet series

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s}$$

[2] In fact, more is true. The infinite series $\sum_{n=1}^{\infty} \frac{f(n)}{n^s}$ is then absolutely convergent for any complex number in the right half plane $\Re(s) > s_0$ and defines there a holomorphic function. However, we shall not make use of these notions.

is convergent for $s > 1$. This is quickly checked for example using that, for $n \geq 2$, one has $\frac{1}{n^s} \leq \int_{n-1}^{n} x^{-s}\,dx$, and therefore, for $s > 1$

$$\sum_{n\geq 2} \frac{1}{n^s} \leq \int_1^\infty x^{-s}\,dx = s - 1.$$

Clearly $D_C(s) = \zeta(s)$ and $D_{C_k}(s) = \zeta(s-k)$, where $C_k : n \mapsto n^k$. Using these identities and the homomorphism property of $f \mapsto D_f$ we obtain

$$\begin{aligned} D_{\sigma_k}(s) &= D_{C*C_k}(s) = \zeta(s-k)\zeta(s) \\ D_\mu(s) &= 1/\zeta(s) \\ D_\varphi(s) &= D_{C_1/\mu} = \zeta(s-1)/\zeta(s) \\ D_\lambda(s) &= \zeta(2s)/\zeta(s). \end{aligned}$$

For the last formula we used $C * \lambda = \chi_\square$, where $\chi_\square(n) = 1$ if n is a perfect square, and $\chi_\square(n) = 0$ otherwise.

If f is multiplicative, then the Euler product, interpreted as (infinite) product of functions, also converges absolutely towards $D_f(s)$ if the series defining $D_f(s)$ converges absolutely. In this way, we obtain for example Euler's famous identity

$$\sum_{n\geq 1} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}} \qquad (s > 1).$$

Chapter 3

Diophantine equations

6. Introduction

A diophantine equation is an equation of the form

$$f(x_1, \ldots, x_n) = 0,$$

where $f(a_1, \ldots, x_n)$ is a polynomial in a number n of unknowns with integer coefficients, and where we seek for integral or rational solutions. Many classical problems lead to questions for the solubility or a description of all solutions of diophantine equations.

Example. We want to determine all right triangles whose side lengths are integral or rational. Pythagoras's theorem tells us that our problem is equivalent to solving the diophantine equation

$$a^2 + b^2 = c^2$$

(i.e. the diophantine equation $f(a, b, c) := a^2 + b^2 - c^2 = 0$). The integral solutions of this equation are called *Pythagorean triples.* We shall describe them all in later sections.

Example. We want to determine all natural numbers $n \in \mathbb{Z}_{\geq 1}$ which occur as the area of a right triangle with rational side lengths. Such integers are called *congruent numbers.* By basic theorems from Euclidean geometry a positive integer n is a congruent number if the following system of diophantine equations is solvable in rational numbers a, b, c:

$$n = \frac{ab}{2}, \quad a^2 + b^2 = c^2.$$

(Note that if this equation has a solution then it has also a solution with a, b, c all three positive).

One can always reduce a system of diophantine equations like the previous one to a single equation. Namely, the system of equations $f_1 = f_2 = \cdots = f_r = 0$ has obviously the same rational or integral solutions as the single equation $f_1^2 + f_2^2 + \cdots + f_r^2 = 0$. However, this is more a theoretical remark. In practice there are better methods to treat systems of diophantine equations. In our case the reader can for example eliminate the variable b by using that $b = 2n/a$ and requiring that the resulting equation $a^4 + 4n^2 = a^2c^2$ has to be solved in rational number a, c with $a \neq 0$. We shall also come back to the congruent number problem in later chapters.

In the following we shall discuss various types of diophantine equations. However before going into details we would like to emphesize that diophantine equations belong to the very heart of mathematics. For the student who encounters them first they might seem to be merely a challenging exercise. However this is far off the truth. We indicate two aspects of this in the following two subsections.

6.1. Hilbert's tenth problem. In a sense every subset of objects in a given set of countably many objects which can be enumerated by some effective procedure (like the subset of prime numbers in the set of all positive integers or the subset of solvable groups in the set of all finite groups) can be encoded by a suitable family of diophantine equations. Thus a huge part of mathematics or computer science is equivalent to the question of solubility of diophantine equations. This is the philosophical interpretation of Matiyasevich's Theorem. This theorem answers also the tenth of Hilbert's 23 problems which Hilbert proposed on the second International Congress of Mathematics, which took place 1900 in Paris, as the outstanding mathematical problems of the coming century.

Hilbert's tenth Problem. *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.*

The notion of *process* as Hilbert called it or, as we say today, *algorithm* is meanwhile, after a huge amount of work in fundamental research in mathematics during the first part of the 20th century, rather well understood. This research is connected to names like Gödel, Turing, Neumann, Church and many others[1].

In particular, we have a mathematical precise notion of *recursive sets*. These are those subsets B of the set of non-negative integers $\mathbb{Z}_{\geq 0}$ for which there exists an algorithm which decides for a given integer $n \geq 0$ whether it belongs to B or not. The notion algorithm, and thus recursive set, has been defined in many different ways. For instance we might define B to be recursive by requiring that we can write a program in our favorite programming language which takes as input an integer $n \geq 0$ and outputs True if n belongs to B and False otherwise. However, all definitions have been proven to describe exactly the same family of subsets of $\mathbb{Z}_{\geq 0}$. This led in the end to what is called *Church-Turing thesis*: A function $\mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ (like the characteristic function of a set B) is computable by a human being ignoring resource limitations if and only if it is computable by a Turing machine.

A related notion is the notion of a *recursive enumerable set*. These are subsets A of $\mathbb{Z}_{\geq 0}$ for which there is an algorithm (e.g. a Turing machine) which enumerates A. We might think of such a set as semidecidable in the sense that, given an integer $n \geq 0$, we will know after running our algorithm and waiting long enough that n is in A once it is output by the algorithm; on the other hand if we do not get an answer after a certain time we can never be sure that n is not in A. A set B is recursive if and only if B and the complement $\mathbb{Z}_{\geq 0} \setminus B$ are both recursively enumerable.

[1] These mathematical-philosophical considerations which led amongst others to the exact notion of algorithm are in a certain sense the basis of the social changes caused by the increasing digitization of information and automatizing of procedures. On the other hand, fundamental research is also not independent of social changes.

Back to diophantine equations we define

Definition (Diophantine Set)**.** A subset $A \subset \mathbb{Z}_{\geq 0}^n$ is called *diophantine*, if there exists a polynomial $f(x_1, \dots, x_r, n)$ with integer coefficients such that

$$A = \{n \in \mathbb{Z}_{\geq 0} : \exists x_1, \dots, x_r \in \mathbb{Z}^r \colon f(x_1, \dots, x_r, n) = 0\} ..$$

Sometimes in the definition of diophantine sets the quantification over all integers is replaced by a quantification over all non-negative integers. However, these two definitions are equivalent. Namely, if a polynomial $f(x_1, \dots, x_r)$ is given we can easily construct a polynomial g such that f is solvable in non-negative integers if and only g is solvable in integers: one can take $g(x_1, \dots, x_r) = \prod_\varepsilon f(\varepsilon_1 x_1, \dots, \varepsilon_r x_r)$, where ε runs through all vectors of length r with ±1 as entries. Vice versa, if we set

$$\begin{aligned} &g(y_1, z_1, v_1, w_1 \dots, y_r, z_r, v_r, w_r) \\ &\qquad\qquad = f(y_1^2 + z_1^2 + v_1^2 + w_1^2, \dots, y_r^2 + z_r^2 + v_r^2 + w_r^2), \end{aligned}$$

then f is solvable in integers if and only if g is solvable n non-negative integers. For proving latter equivalence we cite Lagrange's four-square theorem, which states that every non-negative integer can be written as sum of four squares.

It is easy to write a computer program which enumerates a diophantine set. For example we can write a program which tries all possible values for n, x_1,..., x_r, in increasing order of the sum of their absolute values, and prints n whenever $f(x_1, ..., x_r, n) = 0$.However, much more involved is the prove of the following theorem.

Theorem (Yuri Matiyasevich, 1970)**.** *Every recursive enumerable set is diophantine.*

We encountered already a diophantine set, namely the set of congruent numbers. It is easy to find other examples. Matiyasevich's theorem gives a philosophical reason for this. The reader might want to look up a description of the set of prime numbers as diophantine set. Matiyasevich provides in fact an answer to Hilbert's tenth problem.

Corollary (Solution of Hilbert's tenth problem). *There is no algorithm which decides, for a given diophantine equation, whether it is solvable or not.*

For deriving the corollary let A be a recursive enumerable but not recursive set. Since A is diophantine we can describe it by a polynomial $f(x_1, \ldots, x_r, n)$ as in the definition of diophantine sets. Since A is not recursive there exists no algorithm which decides whether a given n is in A, i.e. whether, for a given n, the diophantine equation $f(x_1, \ldots, x_r, n) = 0$ is solvable in integers x_1,..., x_r. However, it is not a priori clear that there exist recursive enumerable which are not recursive. In mathematical logic it is shown that this is indeed the case.

6.2. Relations to geometry. The second indication for the importance of diophantine equations is that they relate arithmetic and geometry in a deep way which is by far not yet fully explored. If we admit as solutions of the equation $f(x_1, \ldots, x_n) = 0$ real numbers we obtain a geometrical object: a hyper-surface in the affine space $\mathbb{R}^n$. As example consider the equation $x^2 + y^2 = 1$, which defines over the reals the unit circle in the affine plane. It is wise to admit even complex numbers as solutions: we then obtain complex algebraic varieties and we are in the heart of complex algebraic geometry. We shall see that in the case of diophantine equations which define complex curves the topological properties of these curves already determine the qualitative behavior of the set of solutions of the underlying diophantine equation. Finer information is provided by studying in addition the congruences $f(x_1, \cdots, x_n) \equiv 0 \bmod p$ for primes p. We shall get a glimpse of this when discussing Legendre's theorem.

7. Diophantine equations in one variable

Consider a polynomial in one variable with integer coefficients

$$P(x) = a_n x^n + \cdots + a_1 x + a_0.$$

We want to determine all rational solutions of $P(x) = 0$ if there are any. In other words, we look for all integers r and $s \geq 1$ such that $P(r/s) = 0$. We can assume that the fraction r/s is given in its lowest terms, i.e. that $\gcd(r,s)=1$. We can moreover assume that a_n and a_0 are different from 0 (otherwise omit the terms which are zero and divide by a suitable power of x).

Writing out the equation $P(r/s)=0$ and multiplying by s^n we obtain

$$a_n r^n + a_{n-1} r^{n-1} s \cdots + a_1 r s^{n-1} + a_0 s^n = 0.$$

But then s divides $a_n r^n$, and since $\gcd(r,s)=1$, it divides even r. Similarly we note that r divides a_0. We thus have proved

Theorem. *One has*

$$\{r/s \in \mathbb{Q} : \gcd(r,s)=1,\ s \geq 1,\ P\left(r/s\right)=0\} \\ \subseteq \left\{r/s \in \mathbb{Q} : r \mid a_0, s \mid a_n,\ s \geq 1\right\}.$$

We have therefore reduced the problem of solving $P(x)=0$ in rational numbers to the computation of the values $P(r/s)$ for the finitely many rational numbers r/s described in the theorem. There is an interesting consequence worth to be noted.

Corollary. *Assume that P is monic (i.e. that $a_n=1$). Then every rational solution of $P(x)=0$ is integral.*

We note a special case of this:

Corollary. *Let n be a positive integer which is not a perfect square. Then $\sqrt{n}$ is irrational.*

Proof. Indeed, $\sqrt{n}$ is a solution of $x^2-n=0$. If it were rational than it would be integral, thus n a perfect square. □

Corollary. *$\sqrt{2}$ is irrational.*

8. Linear diophantine equations

Next we look at diophantine equations in several variable but restrict the degree. We start with the degree 1 case, in other words we want to solve an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b$$

in integers x_j. The coefficients a_j and b are as usual assumed to be integers and the a_j are not all zero. We could also look for rational solutions, but this problem is quickly solved. If, say, a_n is different from zero, then the set of solutions in rational numbers equals the the set of all vectors of the form

$$(x_1, \ldots, x_{n-1}, (b - a_1x_1 + \cdots + a_{n-1}x_{n-1})/a_n),$$

where the x_j are arbitrary rational numbers. If we ask for integral solutions the answer is not so obvious. The next theorem tell us that we can restrict to such equations where the a_j $(j = 1, 2, \ldots, n)$ are relatively prime.

Theorem. *The diophantine equation $a_1x_1 + \cdots + a_nx_n = b$ possesses a solution in integers if and only if* $\gcd(a_1, \ldots, a_n)$ *divides b.*

The theorem was already proved in Section 1. Namely, the given equation has integer solutions if and only if b is contained in the ideal generated by the numbers a_j, and we have seen that is ideal is generated by the gcd of the a_j.

So we assume that our diophantine equation has solutions. We can then divide the equation by the gcd of the a_j, and the resulting equation has the property that the a_j are relatively prime, which we assume from now on. The question remains how to find and describe the solutions.

Let us assume for a moment that $n = 2$. We have seen in Section 1 that the extended Euclidean algorithm gives generates a solution $x_1^{(0)}$, $x_2^{(0)}$ of $a_1x_1 + a_2x_2 = b$. It is now easy to obtain from this all solutions.

Theorem. *The integral solutions of the equation $a_1x_1 + a_2x_2 = b$ are given by*

$$x_1 = x_1^{(0)} - ta_2, \quad x_2 = x_2^{(0)} + ta_1,$$

where t runs through the integers.

Proof. If x_1 and x_2 are solutions then $a_1(x_1 - x_1^{(0)}) + a_2(x_2 - x_2^{(0)}) = 0$. It follows that a_2 divides $a_1(x_1 - x_1^{(0)})$, and since a_1 and a_2 are relatively prime, that a_2 divides in fact $x_1 - x_1^{(0)}$. Therefore $x_1 - x_1^{(0)} = aa_2$ for some integer t. Similarly, $x_2 - x_2^{(0)} = ta_1$ for some integer u. From $a_1(ta_2) + a_2(ua_1) = 0$ we obtain $s = -t$. Vice versa it is clear that any pair x_1, x_2 of the given form provides a solution. $\square$

For extending the last theorem to more tan two variables it is useful to reformulate it in a slightly different form. For this we can assume that $x_1^{(0)} = bu_1$, $x_2^{(0)} = bu_2$ with integers u_1, u_2 satisfying $a_1u_1 + a_2u_2 = 1$. We can write then the general solution of $a_1x_1 + a_2x_2 = b$ in the form

$$(x_1, x_2) = (b, t)\begin{bmatrix} u_1 & u_2 \\ -a_2 & a_1 \end{bmatrix}.$$

The matrix has determinant 1, its inverse is $\begin{bmatrix} a_1 & -u_2 \\ a_2 & u_1 \end{bmatrix}$. Vice versa one verifies that for any such matrix U, the vector $(b, t)U^{-1}$ runs through all solutions of our equation when t runs through $\mathbb{Z}$. This is in fact true for any number of variables.

Theorem. *Let U be an $n \times n$-matrix of determinant ± 1 and with $(a_1, \ldots, a_n)'$ as first column*[2]*. Then a vector $(x_1, \ldots, x_n)$ is an integral solution of the equation $a_1x_1 + \cdots + a_nx_n = b$ if and only if it is of the form*

$$(x_1, \ldots, x_n) = (b, t_2, \ldots, t_n)U^{-1},$$

where $t_2, \ldots, t_n$ are integers.

Proof. If $(x_1, \ldots, x_n)$ is of the given form then the x_j are integers. For this we have to verify that U^{-1} has integers as entries. But this follows from the formula $U^{-1} = \det(U)^{-1}U^*$, where U^* is the adjunct of U. Thus, if $(x_1, \ldots, x_n)$ is of the given form it has integral entries and satisfies $(x_1, \ldots, x_n)U = (b, t_2, \ldots, t_n)$, in particular, $(x_1, \ldots, x_n)$ multiplied with the first row of U equals b. But this product is by assumption nothing else but $a_1x_1 + \cdots + a_nx_n$.

Vice versa, if $(x_1, \ldots, x_n)$ is a solution, then $(x_1, \ldots, x_n)U = (b, t_2, \ldots, t_n)$ for suitable integers t_j. □

It remains the question whether such a matrix as in the theorem always exists and how to compute it. For this we note that the $n \times n$-matrices with determinant ± 1 and integral entries form a group, which is usually denoted by $\mathrm{GL}(n, \mathbb{Z})$ with respect to matrix multiplication. It is not important to understand the last statement.

[2] For a vector or matrix X we use X' for the transpose of X.

It suffices to know that for every two matrices U and V in $\mathrm{GL}(n, \mathbb{Z})$ their product and their inverses are in $\mathrm{GL}(n, \mathbb{Z})$ too.

Theorem. *Let a be an integral primitive*[3] *column vector of length n. Then there exists a matrix U in $\mathrm{GL}(n, \mathbb{Z})$ whose first column equals a.*

Proof. The proof will provide also an algorithm for obtaining such a matrix U. In fact we apply the generalized Euclidean algorithm to a for obtaining the desired U.

The extended Euclidean algorithm as explained in Section **??** generates a sequence of vectors $a_0 = a$, a_1, $\ldots$, $a_k = (1, 0, \ldots, 0)'$ starting with a and ending with $(1, 0, \ldots, 0)'$. Each a_{j+1} is obtained from a_j by applying one of the three operations:

(1) Exchange two entries.

(2) Multiply an entry by -1.

(3) Replace the kth entry by the rest after Euclidean division by the lth entry.

But each of these operations corresponds to a matrix multiplication from the left: (1) to multiplication by a permutation matrix, (2) by the diagonal matrix whose diagonal entries are all 1 except for one which is -1, and (3) to a matrix of the form $E + tE_{kl}$, where E is the unit matrix, t an integer and E_{kl} the matrix which has a 1 at the k, lth place and 0 at all others. All these matrices are in $\mathrm{GL}(n, \mathbb{Z})$. Thus we have $a_{j+1} = V_j a_j$ for some V_j in $\mathrm{GL}(n, \mathbb{Z})$, and so $(1, 0, \ldots, 0)' = V_{k-1} \cdots V_1 V_0 a$. The matrix $U = (V_{k-1} \cdots V_1 V_0)^{-1}$ satisfies then $U(1, 0, \ldots, 0)' = a$, i.e. it has a as first column. □

It is not hard to transform this algorithm into an algorithm.

Algorithm: Computation of the matrix U

Pending Exercise

[3] An integral vector is called *primitive* if its entries are relatively prime.

Example. We end this section by an example: we want to determine all solutions of $3x + 5y + 7z = 1$. For this we guess a matrix U in $\mathrm{GL}(n, \mathbb{Z})$ whose first columns equals $(3, 5, 7)$. On can take for example

$$U = \begin{bmatrix} 3 & 1 & 0 \\ 5 & 2 & 0 \\ 7 & 0 & 1 \end{bmatrix}$$

The general integral solution of our equation is therefor

$$\begin{aligned} (x_1, x_2, x_3) &= (1, t, u) \begin{bmatrix} 2 & -1 & 0 \\ -5 & 3 & 0 \\ -14 & 7 & 1 \end{bmatrix} \\ &= (2 - 5t - 17u, -1 + 3t + 7u, u), \end{aligned}$$

where t and u are integers.

9. Special quadratic diophantine equations

In this section we discuss diophantine equations of the form $f(x, y) = 0$, where f is a polynomial of degree 2. Recall that this this means that f is a linea combination of monomials $x^k y^l$ with $k + l \leq 2$ and that we have equality for at least one monomial. In fact we shall restrict here to special f, namely $f(x, y) = x^+ y^2 - 1$, and the family $f_n = x^2 - ny^2 - 1$ (n a poitive integer). A complete theory for arbitrary f of degree 2 can be found in Chapter **??** on conic sections. In this section we shall also encounter for the first time the use of geometric arguments to solve diophantine problems.

9.1. Rational points on the unit circle. We consider the diophantine equation

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

The set of solutions in real numbers is the unit circle in the Euclidean plane. The set of integral solutions consists merely of the four points $(\pm 1, 0)$, $(0, \pm 1)$. However, if we ask for rational solutions the question becomes much more interesting: there are plenty such solutions like e.g. $(3/5, 4/5)$, $(15/17, 8/17)$. In fact, there are infinitely many solutions. The idea to find them all is already describes in a famous book on diophantine problems which is attributed to Diophant who lived around 250 AD.

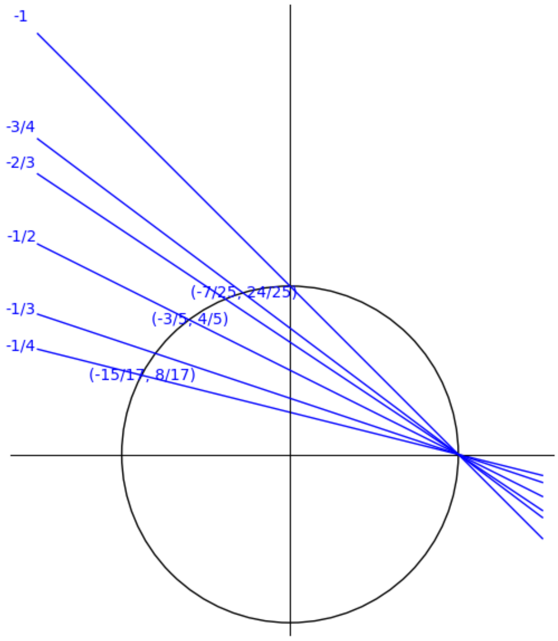


Figure 1. Diophant's method of parametrizing the points on the unit circle by the lines of a pencil.

The idea is to consider the pencil through the point $(1, 0)$ on the unit circle, i.e. the set of lines through this point; see Fig. 1. It is clear that every line in the plane intersect the unit circle in at most two points. A line intersects the unit circle in only one point if it is a tangent. A line in our pencil through $(1, 0)$ which is different from the tangent line (the line perpendicular to the x-axis) intersect the unit circle in exactly one point different from $(1, 0)$. Vice versa, every point different from $(1, 0)$ lies on exactly one line in our pencil. In other words, the application which associates to every line of the pencil its second intersection point with the unit circle defines a bijection.

But the lines of our pencil different from the tangent are given by the equations $y = \lambda(x-1)$, where λ runs trough the reals. As we shall verify in a moment, the slope λ is rational if and only if (x, y) has rational entries. This will therefore solve our diophantine problems.

The intersection point of $y = \lambda(x-1)$ with the unit circle is quickly computed: Elimination of the variable y gives

$$\begin{aligned} &x^2 + \lambda^2(x-1)^2 = 1 \quad (x \neq 1),\\ &x + 1 + \lambda^2(x-1) = 0,\\ &x = \frac{\lambda^2 - 1}{\lambda^2 + 1}, \end{aligned}$$

and inserting back this expression for x in $y = \lambda(x-1)$ then $y = \frac{-2\lambda}{\lambda^2+1}$.

Clearly, if λ is rational so are x and y. Vice versa, if x, y are rational the formula $\lambda = \frac{y}{x-1}$ shows that λ is in $\mathbb{Q}$. We can summarize our reasoning by the following theorem.

Theorem. *One has*

$$\begin{aligned} &\left\{(x,y) \in \mathbb{Q}^2 \colon x^2 + y^2 = 1, (x,y) \neq (1,0)\right\}\\ &\qquad\qquad = \left\{\left(\tfrac{\lambda^2-1}{\lambda^2+1}, \tfrac{-2\lambda}{\lambda^2+1}\right) : \lambda \in \mathbb{Q}\right\} \end{aligned}$$

Diophant's method can obviously also applied to other quadratic diophantine equations. The reader might try to apply it for example to the diophantine equation $x^2 + xy + y^2 = 0$. However, Diophant's method depends on the knowledge of at least one solution for basing the pencil on it. Such a point does not necessarily exist. In Section **??** we shall give an answer to the question when a solution exist and when not, and how to compute one.

If we look again at our derivation of the equality of the theorem we see that we could apply this method also to other fields, not only the rational numbers. In particular, we can apply it to the finite fields $\mathbb{Z}/p\mathbb{Z}$, where p is a prime.