3. An outline of the proof of Roth's Theorem

$$N'(m) \approx c_2(F)\, m^{2/d}$$

if either $d = 3$ and the discriminant of F is not of some rather special type, or if $F(x, y) = x^d + y^d$ for some $d \geqq 3$. To generalize these results to arbitrary forms $F(x, y)$ appears to be extremely difficult.

The methods of Thue, Siegel and Roth do not enable one to find bounds for the size $|x| + |y|$ of solutions of Thue's equation, and hence they provide no method to find all the solutions of such an equation. Therefore these methods are called " non-effective ". Effective results will be discussed in §5.

3. An outline of the proof of Roth's Theorem

3.1. We shall follow Cassel's rearrangement (Cassels (1957), ch. VI) of Roth's proof. It is easy to see that we may restrict ourselves to the case when α is an algebraic *integer* of degree $d > 1$.

Suppose we tried to modify the proof of Liouville's Theorem as follows. In step (a) we pick a polynomial $P(x)$ with rational integer coefficients which has a root at α of order i and which has degree r. Next, in step (b) we suppose that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{\mu}}, \tag{3.1}$$

and Taylor's expansion

$$P\left(\frac{p}{q}\right) = \sum_{j=i}^{r} \left(\frac{p}{q} - \alpha\right)^j \frac{1}{j!} P^{(j)}(\alpha)$$

yields $\left|P\left(\frac{p}{q}\right)\right| \leqq cq^{-\mu i}$. Finally (c) we have $P\left(\frac{p}{q}\right) \neq 0$ whence $\left|P\left(\frac{p}{q}\right)\right| \geqq q^{-r}$ for all but finitely many rationals $\frac{p}{q}$. Hence if (3.1) has infinitely many solutions, then $\mu i \leqq r$ or

$$\mu \leqq \left(\frac{i}{r}\right)^{-1}.$$

Hence one should try to make $\frac{i}{r}$ as large as possible. But it is clear that

always $\frac{i}{r} \leqq \frac{1}{d}$, and that $\frac{i}{r} = \frac{1}{d}$ if $P(x)$ is a power of the defining polynomial of α. Hence this method only gives $\mu \leqq d$, i.e. nothing better than Liouville's result.

3.2. In order to improve on this estimate, Thue and Siegel use a polynomial $P(x_1, x_2)$ in two variables, and Schneider (1936) and Roth use a polynomial $P(x_1, \ldots, x_m)$ in many variables. It is necessary to define the order of vanishing of $P(x_1, \ldots, x_m)$ at a given point $(\xi_1, \ldots, \xi_m)$. The simplest definition would be to take the smallest value of $i_1 + \ldots + i_m$ for which the mixed partial derivative

$$P^{(i_1,\ldots,i_m)}(\xi_1, \ldots, \xi_m) \neq 0 . \tag{3.2}$$

But it is necessary to study polynomials $P(x_1, \ldots, x_m)$ which have rather different degrees in $x_1, \ldots, x_m$, and hence it will be better to attach different weights to the integers $i_1, \ldots, i_m$ in (3.2). Thus Roth defines the *index of* P *at* $(\xi_1, \ldots, \xi_m)$ with respect to a given m-tuple of positive integers $(r_1, \ldots, r_m)$ as the least value of

$$\frac{i_1}{r_1} + \ldots + \frac{i_m}{r_m}$$

for which (3.2) holds, if $P \not\equiv 0$, and as $+\infty$ if $P \equiv 0$.

3.3. The steps (a), (b), (c) in the proof of Liouville's Theorem are now replaced by new steps (a), (b), (c).

(a) LEMMA 3A. *Suppose α is an algebraic integer of degree $d > 1$. Suppose $\varepsilon > 0$ and m is an integer with*

$$m > 8d^2 \varepsilon^{-2} . \tag{3.3}$$

Let $r_1, \ldots, r_m$ be positive integers. Then there is a polynomial $P(x_1, \ldots, x_m) \not\equiv 0$ with rational integer coefficients such that

(i) *P has degree at most r_h in x_h $(h = 1, \ldots, m)$.*

(ii) *P has index at least $\frac{m}{2}(1-\varepsilon)$ at $(\alpha, \ldots, \alpha)$ with respect to $(r_1, \ldots, r_m)$.*

(iii) *$H(P) \leqq B^{r_1+\ldots+r_m}$ where $B = B(\alpha)$.*

Here $H(P)$ is the *height* of P, i.e. the maximum of the absolute values

of its coefficients. By virtue of (ii) the average of $\frac{i_h}{r_h}(h=1, \ldots, m)$ when $P^{(i_1,\ldots,i_m)}(\alpha, \ldots, \alpha) \neq 0$ is at least $\frac{1}{2}(1-\varepsilon)$, which is rather better than $\frac{i}{r} = \frac{1}{d}$ we had in the proof of Liouville's Theorem.

To prove the lemma we put

$$P(x_1, \ldots, x_m) = \sum_{j_1=0}^{r_1} \ldots \sum_{j_m=0}^{r_m} C(j_1, \ldots, j_m)\, x_1^{j_1} \ldots x_m^{j_m}.$$

The $N = (r_1+1) \ldots (r_m+1)$ coefficients $C(j_1, \ldots, j_m)$ are unknown integers we have to determine such that (ii) and (iii) hold. The condition (ii) means that

$$P^{(i_1,\ldots,i_m)}(\alpha, \ldots, \alpha) = 0 \tag{3.4}$$

whenever

$$\frac{i_1}{r_1} + \ldots + \frac{i_m}{r_m} < \frac{m}{2}(1-\varepsilon). \tag{3.5}$$

Since (3.4) is always true if $i_h > r_h$ for some h, the number of non-trivial equations (3.4) is the number of points $\left(\frac{i_1}{r_1}, \ldots, \frac{i_m}{r_m}\right)$ in the unit cube $(0 \leqq \xi_1 \leqq 1, \ldots, 0 \leqq \xi_m \leqq 1)$ with (3.5). One can show that

$$\frac{\text{The number of points in the cube with (3.5)}}{N = \text{the total number of points in the cube}} \to 0 \tag{3.6}$$

as $m \to \infty$, independently of $r_1, \ldots, r_m$. This is just the law of large numbers in probability theory, since the " independent variables " $i_1/r_1, \ldots, i_m/r_m$ each have expectation value $\frac{1}{2}$. In fact an appeal to probability theory is not necessary and a simple combinatorial argument shows that the left hand side of (3.6) is at most $2^{1/2}m^{-1/2}\varepsilon^{-1}$, and hence by (3.3) is at most $1/(2d)$. Thus the number of non-trivial conditions (3.4) is at most $N/(2d)$. Each condition (3.4) is a homogeneous linear equation in the unknowns $C(j_1, \ldots, j_m)$ with coefficients in the field $\mathbf{Q}(\alpha)$. (I.e. the field obtained by adjoining α to the field $\mathbf{Q}$ of rationals). Hence each condition follows from d linear homogeneous equations whose coefficients are rational integers. Hence altogether our unknown integers $C(j_1, \ldots, j_m)$ have to satisfy

at most $N/2$ linear homogeneous equations with rational integer coefficients. But it is known that any system of linear homogeneous equations with rational integer coefficients where the number of equations is at most $\frac{1}{2}$ times the number of unknowns has a non-trivial solution in rational integers which are bounded in terms of the size of the coefficients. Carrying out all the estimates one sees that (iii) can be satisfied in addition to (ii).

3.4. We now turn to step

(b) It can be shown that if $\frac{p_1}{q_1}, \ldots, \frac{p_m}{p_m}$ are solutions of $\left|\alpha - \frac{p}{q}\right| < q^{-2-\delta}$ and if some further conditions are satisfied, then $P\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) = 0$. In fact a little more is true:

LEMMA 3B. *Suppose* $\left|\alpha - \frac{p_h}{q_h}\right| < q_h^{-2-\delta}$ $(h=1, 2, \ldots, m)$ *where* $0 < \delta < \frac{1}{12}$. *Suppose* $0 < \varepsilon < \delta/20$ *and suppose that* $q_h \geqq c_0(\alpha, \delta)$ $(h = 1, \ldots, m)$ *and*

$$(3.7) \qquad r_1 \log q_1 \leqq r_h \log q_h \leqq (1+\varepsilon) r_1 \log q_1 \qquad (h=1, \ldots, m).$$

Now if all the conditions of Lemma 3A *are satisfied and if P is the polynomial of that lemma, then the index of P at* $\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right)$ *with respect to* $(r_1, \ldots, r_m)$ *is* $\geqq \varepsilon m$.

To prove this lemma we shall use Taylor's formula:

$$P\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) = \sum_{i_1=0}^{r_1} \cdots \sum_{i_m=0}^{r_m} \left(\frac{p_1}{q_1} - \alpha\right)^{i_1} \cdots \left(\frac{p_m}{q_m} - \alpha\right)^{i_m} \frac{P^{(i_1, \ldots, i_m)}(\alpha, \ldots, \alpha)}{i_1! \ldots i_m!}.$$

By (ii) of Lemma 3A only terms with $\frac{i_1}{r_1} + \ldots + \frac{i_m}{r_m} \geqq \frac{m}{2}(1-\varepsilon)$ can be non-zero. For these terms we have

$$\left|\left(\frac{p_1}{q_1}-\alpha\right)^{i_1}\ldots\left(\frac{p_m}{q_m}-\alpha\right)^{i_m}\right| \leqq q_1^{-(2+\delta)i_1}\ldots q_m^{-(2+\delta)i_m}$$
$$= \left(q_1^{r_1(i_1/r_1)}\ldots q_m^{r_m(i_m/r_m)}\right)^{-2-\delta}$$
$$\leqq q_1^{-r_1(2+\delta)((i_1/r_1)+\ldots+(i_m/r_m))} \leqq q_1^{-r_1(2+\delta)\frac{1}{2}m(1-\varepsilon)}$$
$$< \left(q_1^{r_1}\ldots q_m^{r_m}\right)^{-\frac{1}{2}(2+\delta)(1-\varepsilon)/(1+\varepsilon)} < \left(q_1^{r_1}\ldots q_m^{r_m}\right)^{-(1+(\delta/4))}$$

by (3.7) and since $0 < \varepsilon < \delta/20$. Using this estimate as well as part (iii) of Lemma 3A it is not hard to show that

$$\left|P\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right)\right| < \left(q_1^{r_1}\ldots q_m^{r_m}\right)^{-1}$$

if $q_h \geqq c_0(\alpha, \delta)$ $(h=1, \ldots, m)$, hence that

$$P\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right) = 0\,.$$

Thus the index of P at $\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right)$ is positive, and a slight extension of the argument shows that the index with respect to $(r_1, \ldots, r_m)$ is at least εm.

3.5. Finally we turn to step

(c) If one could show that $P\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right) \neq 0$, then this would contradict Lemma 3B, and this contradiction would show that the inequality (2.2) has only finitely many solutions. But to show that $P\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right) \neq 0$ was very easy for $m = 1$ and it is rather difficult when $m > 1$. To get a contradiction to Lemma 3B it will suffice to show that the index of P at $\left(\frac{p_1}{q_1},\ldots,\frac{p_m}{q_m}\right)$ with respect to $(r_1, \ldots, r_m)$ is less than εm. When $m = 2$ the situation is a little simpler than in the general case. Siegel (1921a) devised an algebraic argument to deal with this case, and Schneider (1936) devised a more general arithmetical argument. The latter argument was considerably sharpened by Roth. The following lemma of Roth is called Roth's Lemma.

LEMMA 3C. *Suppose* $0 < \varepsilon < 1/12$ *and let* m *be a positive integer. Put* $\omega = 24\cdot 2^{-m}(\varepsilon/12)^{2m-1}$. *Let* $r_1, \ldots, r_m$ *be positive integers with*

(3.8) $$\omega r_h \geqq r_{h+1} \qquad (h=1, \ldots, m-1)\,.$$

Let $\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}$ *be rationals in their lowest terms with positive denominators and with* $q_h^{\omega} \geqq 2^{3m}$ *and* $q_h^{r_h} \geqq q_1^{r_1}$. *Further let* $P(x_1, \ldots, x_m)$ *be a polynomial with rational integer coefficients, not identically zero, of degree* $\leqq r_h$ *in* x_h $(h=1, \ldots, m)$ *and with* $H(P) \leqq q_1^{\omega r_1}$. *Then the index of* P *at* $\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right)$ *with respect to* $(r_1, \ldots, r_m)$ *is* $\leqq \varepsilon$.

The proof of this lemma is ingenious and complicated and will not be given here. It uses " generalized Wronskians ", i.e. determinants whose entries are mixed partial derivatives of certain polynomials. Some condition like (3.8) is necessary, for otherwise if $m = 2$, say, the polynomial $P(x_1, x_2) = (x_1 - x_2)^r$ would have an index as large as 1 at every point (ξ, ξ).

The lemma is proved by induction on m. Only the case $m = 1$ is simple and will be proved here. Suppose $P(x)$ has a zero of order l at p/q. Then

$$P(x) = (qx-p)^l R(x)$$

where $R(x)$ has rational integer coefficients by Gauss' Lemma. We have

$$q^l \leqq H(P) \leqq q^{\omega r_1} = q^{\varepsilon r_1}$$

(since $\omega = \varepsilon$ when $m=1$), whence $l/r_1 \leqq \varepsilon$. But l/r_1 is the index of P at $\frac{p}{q}$ with respect to (r_1).

Now if there are infinitely many rationals $\frac{p}{q}$ with (2.2), then both Lemma 3B and Lemma 3C can be satisfied. (One picks $0 < \delta < 1/12$, then $0 < \varepsilon < \delta/20$, then m with (3.3), then rationals $\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}$ with (2.2) and with rapidly increasing denominators, and finally one picks $r_1, \ldots, r_m$). These two lemmas together give the desired contradiction, and Roth's Theorem follows.

3.6. The reason why this proof is non-effective is that one needs m very good approximations to α rather than just one, in order to get the desired contradiction. For Thue's and Siegel's Theorems one needs two such approximations. In fact Davenport (1968) found a function $\kappa_0(d)$

defined for $d = 3, 4, \ldots$ with $\kappa_0(3) = 1 + \sqrt{3}$ and with $|\kappa_0(d) - \frac{1}{2}d| \leqq c_1$ such that for an algebraic number α of degree d and for $\kappa > \kappa_0(d)$ there is a computable constant $c_2 = c_2(\alpha, \kappa)$ such that the inequality $\left|\alpha - \frac{p}{q}\right| < q^{-\kappa}$ has at most one solution p/q with $(p, q) = 1$ and $q > c_2$. But earlier Schinzel (1967) had pointed out that this is true with $\kappa > 3\sqrt{d/2}$ in place of $\kappa > \kappa_0(d)$. If d is large, then $\kappa_0(d) > 3\sqrt{d/2}$, and hence in this case Schinzel's result is better than Davenport's. Earlier Siegel (1937) and Hyyrö (1964) had shown results of this kind for numbers α of the type $\alpha = \sqrt[d]{(a/b)}$, or rather for the corresponding Thue's equation $ax^d - by^d = m$.

4. Some generalizations of Roth's Theorem

4.1. In this section we shall discuss several generalizations of Roth's Theorem, but not the generalization to simultaneous approximation, which will be taken up in §7.

The *height* $H(\beta)$ of an algebraic number β is defined by $H(\beta) = H(P)$, where P is the defining polynomial of β. Roth (1955b) enunciated and LeVeque ((1955), vol. 2, ch. 4) proved

Theorem 4A. *Let α be algebraic, K an algebraic number field and $\delta > 0$. There are only finitely many elements β of K with*

$$|\alpha - \beta| < H(\beta)^{-2-\delta}. \tag{4.1}$$

Neither α nor K need to be real here. When K is the field $\mathbf{Q}$ of rationals, then Theorem 4A reduces to Roth's Theorem. Since every number field contains $\mathbf{Q}$ as a subfield, it follows from Dirichlet's Theorem that the number 2 in the exponent in (4.1) is best possible if α is real.

In fact if α and K are real, then the exponent is best possible in a somewhat less trivial sense: Suppose K is a real or complex number field of degree t and β in K has degree d. Then d is a divisor of t. Define the *field height* $H_K(\beta)$ of β by $H_K(\beta) = H(P^{t/d})$, where P is the defining polynomial of β. One can show (LeVeque (1955), vol. 2, ch. 4.2) that $c_1(K)\, H(\beta)^{t/d} \leqq H_K(\beta) \leqq c_2(K)\, H(\beta)^{t/d}$, and hence Theorem 4A remains true a fortiori if (4.1) is replaced by

$$|\alpha - \beta| < H_K(\beta)^{-2-\delta}. \tag{4.2}$$