OVERVIEW OF THE PROOF OF ROTH'S THEOREM

JOSEPH HARRISON

We give a brief overview of the proof of Roth's theorem on diophantine approximations. Everything comes directly from Chapter D. in the book of Hindry and Silverman.

1. Liouville's theorem on diophantine approximation

Theorem 1.1 (Liouville 1844)**.** Let α be an algebraic number of degree $d \geq 2$. If a and $q > 0$ are integers such that

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{q^{d+\varepsilon}}$$

then $q^{\varepsilon} \leq c(\alpha)$.

Proof. Let P be the minimal polynomial of α over $\mathbf{Q}$. Then

$$\begin{aligned} |P(a/q)| &= \left|\sum_{i=1}^{d} \frac{P^{(i)}(\alpha)}{i!}\left(\frac{a}{q} - \alpha\right)^{i}\right| \\ &\leq \left|\alpha - \frac{a}{q}\right| \sum_{i=1}^{d} \frac{|P^{(i)}(\alpha)|}{i!} \\ &= c(\alpha)\left|\alpha - \frac{a}{q}\right|. \end{aligned}$$

On the other hand, there is an integer n so that $P(a/q) = n/q^{d}$. Hence

$$\frac{|n|}{q^{d}} \leq \frac{c(\alpha)}{q^{d+\varepsilon}}.$$

If $q^{\varepsilon} > c(\alpha)$ then $|n| < 1$ so $n = 0$. This is impossible since then $P(a/q) = 0$, but P is irreducible over $\mathbf{Q}$. Hence $q^{\varepsilon} \leq c(\varepsilon)$ as required. **Q.E.D.**

2. Roth's theorem

Theorem 2.1. Let K be a number field, α an algebraic number, $\delta > 0$, and let S be a finite set of places of $K(\alpha)$. Then there are only finitely many $\beta \in K$ satisfying

$$\prod_{v \in S} \min\{1, |\beta - \alpha|_v\} \leq \frac{1}{H_K(\beta)^{2+\delta}}. \tag{1}$$

The steps involved in the proof of Roth's theorem already take place in the proof of Liouville's theorem, albeit in a rather trivial way. It would perhaps be clearer to demonstrate the steps via Thue's theorem (exponent d replaced with $d/2 + 1$ when $d \geq 3$), but in the interest of time we have chosen Liouville's result instead.

(1) Construct an auxiliary polynomial $P \in \mathbf{Z}[X_1, \ldots, X_m]$ that vanishes at α, or possibly $(\alpha, \ldots, \alpha)$. In Liouville's theorem we can take P to be the minimal polynomial of α.
(2) Force P to vanish at approximations of α to high order, e.g., $P(\beta_1, \ldots, \beta_m) = 0$ where each β_i satisfies (1). In the proof of Liouville we only need P to vanish at one approximation.
(3) Order of vanishing of P at $(\beta_1, \ldots, \beta_m)$ is bounded in a way that depends on the heights of the β_i are P. In the proof of Liouville theorem this is provided by the irreducibility of P.
(4) The proof is finished by setting the values of various parameters. This is analogous to choosing $q^\varepsilon \geq c$ in Liouville.

For $(i_1, \ldots, i_m)$ non-negative integers, we define the differential operators

$$\partial_{i_1,\ldots,i_m} P = \frac{1}{i_1! \ldots i_m!} \frac{\partial^{i_1+\cdots+i_m} P}{\partial X_1^{i_1} \ldots \partial X_m^{i_m}}.$$

The Leibniz formula is simpler for these operators and they play better with respect to the norm $|P|_\infty$, which is the maximum absolute value of the coefficients of P.

Definition 2.2. Let $\boldsymbol{\gamma} \in \mathbf{C}^m$ and $r_1, \ldots, r_m$ positive integers. The index with respect to $(\boldsymbol{\gamma}; \boldsymbol{r})$, denoted $\mathrm{Ind}_{(\boldsymbol{\gamma},\boldsymbol{r})}(P)$, is the smallest value of

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

where $\partial_{i_1,\ldots,i_m} P(\boldsymbol{\gamma}) \neq 0$.

Eventually we will take $r_1, \dots, r_m$ to be the degrees of P in each variable. Then

$$0 \leq \mathrm{Ind}(P) \leq m$$

with first equality if and only if $P(\boldsymbol{\gamma}) \neq 0$ and second if and only if $P = c(X_1 - \gamma_1)^{r_1} \dots (X_m)$. This is a normalised notion of order of vanishing that uses the fact that we can make P vanish to high order by increasing the degree.

3. Reductions

Multiply α and a putative approximation by a sufficiently large integer. If the height of the approximation is large we contradict integral Roth.

Proposition 3.1. It suffices to prove Roth's theorem with α an algebraic integer.

Hence we may remove the "min" in the statement of Roth.

Proposition 3.2. Roth's theorem is equivalent to the following statement: Let α be an algebraic integer. For every function $\xi : S \to [0, 1]$ such that $\sum_v \xi_v = 1$ there are finitely many $\beta \in K$ such that

$$|\beta - \alpha|_v \leq \frac{1}{H_K(\beta)^{(2+\delta)\xi_v}}$$

for all $v \in S$.

4. Construction of auxiliary polynomial

If we want to prove Roth with exponent $2 + \delta$ then we take the parameter $0 < \varepsilon < \delta/22$.

Proposition 4.1. Let α be algebraic of degree d over $\mathbf{Q}$. Let m be an integer such that

$$\exp\left(\frac{\varepsilon^2 m}{4}\right) > 2d.$$

Let $r_1, \dots, r_m$ be given. Then there exists $P \in \mathbf{Z}[X_1, \dots, X_m]$ with

(1) $\deg_{X_h}(P) \leq r_h$.
(2) $\mathrm{Ind}_{(\alpha,\dots,\alpha;\boldsymbol{r})}(P) \geq (1 - \varepsilon)m/2$.
(3) $|P|_\infty \leq B^{r_1+\dots+r_m}$ where $B = B(\alpha)$.

The proof involves Siegel's lemma, various height estimates and a counting problem for the $0 \leq i_h \leq r_h$ satisfying

$$\frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \leq (1-\varepsilon)\frac{m}{2}$$

this is where the Gaussian comes from: choosing i_h randomly should make the sum normally distributed and we are asking for the sum to deviate by at least ε from the mean.

5. Large index

We now show that if P vanishes to high order at $(\alpha, \ldots, \alpha)$ and $\beta_1, \ldots, \beta_m$ are very good approximations of α then P vanishes to high order at $\beta_1, \ldots, \beta_m$ too. This is analogous to the Taylor expansion step in Liouville's theorem.

Proposition 5.1. Let $\xi : S \to [0,1]$ be a function with $\sum_v \xi_v = 1$. Suppose $\beta_1, \ldots, \beta_m \in K$ satisfy the following.

(1) (Good approximations to α) for all $h \in [m]$ and $v \in S$

$$|\beta_h - \alpha|_v \leq \frac{1}{H_K(\beta_h)^{(2+\delta)\xi_v}}.$$

(2) (Height control) Each $H_K(\beta_h) \geq C(\alpha, \delta)$ and

$$D := \min\{H_K(\beta_h)^{r_h}\} \leq \max\{H_K(\beta_h)^{r_h}\} \leq D^{1+\varepsilon}.$$

Then

$$\mathrm{Ind}_{(\boldsymbol{\beta},\boldsymbol{r})}(P) \geq \varepsilon m.$$

Let $\theta = \mathrm{Ind}_{(\alpha,\ldots,\alpha;\boldsymbol{r})}(P)$. Given $I = (i_1, \ldots, i_m)$ satisfying $\sum i_j/r_j \leq \theta_0 \leq \varepsilon m$, we need to show that $\partial_{i_1,\ldots,i_m} P(\beta_1, \ldots, \beta_m) = 0$. Using (1) and (2) we show that

$$\prod_{v \in S} |\partial_I P(\beta_1, \ldots, \beta_m)|_v \leq (4H(\alpha))^{d(r_1+\cdots+r_m)} D^{-(2+\delta)(\theta-\theta_0)}$$

by Taylor expansion arguments similar to in Liouville's theorem.

Now we have Liouville's inequality: either $\partial_I P(\boldsymbol{\beta}) = 0$ or

$$\prod_{v \in S} |\partial_I P(\boldsymbol{\beta})|_v \geq \frac{1}{H_K(\partial_I P(\boldsymbol{\beta}))}.$$

If the inequality does hold then we combine this with upper bounds to get an upper bound on $H_K(\beta_h)$ in terms of α, δ. Hence taking C large in the statement ensures $\partial_I P(\boldsymbol{\beta}) = 0$.

6. Small index

This step is analogous to showing that $P(a/q) \neq 0$ in Liouville's theorem. Actually this can't be done for a univariate polynomial, which is the whole reason why we need more variables. Thue used a bivariate polynomial to prove his theorem. At the end we will see that $\eta = \varepsilon/4$ works in the following.

Proposition 6.1 (Roth's lemma)**.** Suppose $P \in K[X_1, \dots, X_m]$ with $\deg_{X_h}(P) \leq r_h$. Let $\boldsymbol{\beta} \in K^m$. Choose $\eta > 0$ satisfying the following conditions.

(1) $r_{h+1}/r_h \leq \eta^{2^{m-1}}$.
(2)

$$\eta^{2^{m-1}} \log \min\{H(\beta_h)^{r_h}\} \geq \log H(P) + 2mr_1.$$

Then

$$\mathrm{Ind}_{(\boldsymbol{\beta},\boldsymbol{r})}(P) \leq 2m\eta.$$

The idea is to find a polynomial $W \in K[X_1, \dots, X_m]$ such that

(1) $h(W) \leq k(h(P) + O(1))$ and $\deg_{X_h}(W) \leq kr_h$ $(k \leq r_m + 1)$.
(2) $W(\boldsymbol{X}) = V(X_1, \dots, X_{m-1})U(X_m)$ so $\mathrm{Ind}(W) = \mathrm{Ind}(V) + \mathrm{Ind}(U)$.
(3)

$$\mathrm{Ind}(W) \geq \frac{k}{2} \min\{\mathrm{Ind}(P), \mathrm{Ind}(P)^2\} - k\frac{r_m}{r_{m-1}}.$$

The proof is then completed by induction on the number of variables.

Let us show how this step is completed in the proof of Thue's theorem. He uses the auxiliary polynomial $P(X, Y) = f(X) + g(X)Y$. Then

$$\partial_{i,0} P(\beta_1, \beta_2) = \partial_i f(\beta_1) + \partial_i g(\beta_1)\beta_2 = 0$$

when $i \leq Ir$. Let

$$W(X) = f\partial g - g\partial f$$

be the Wronskian of (f, g). Then $\partial_k W(\beta_1) = 0$ for $k \leq rI - 1$. Hence $\mathrm{Ind}(W) \geq \mathrm{Ind}(P) - 1/r$. By estimating $|W|_\infty$ we can write down an upper bound for $\mathrm{Ind}(W)$, and hence also for $\mathrm{Ind}(P)$.

The proof of Roth's theorem involves Wronskians in a cleverer way.

7. Completion of the proof

The proof is mostly a case of setting parameters. ε and m have been set, but we also define $\omega = (\varepsilon/4)^{2^{m-1}}$. Since we assume for a contradiction that there are infinitely many good approximations, let us choose one β_1 satisfying

$$H(\beta) \geq \max\left\{C, \exp\left(\frac{m(\log(B) + 2)}{\omega}\right)\right\}.$$

Now $\beta_2, \ldots, \beta_m$ are chosen so that

$$H_K(\beta_{h+1})^\omega \geq H_K(\beta_h)^2.$$

Hence $H_K(\beta_h)$ is growing very quickly. The reason for the ω is that we would like the $H_K(\beta_h)^{r_h}$ to be roughly equal, which leads to

$$r_h = \left\lceil \frac{r_1 \log H_K(\beta_1)}{\log H_K(\beta_h)} \right\rceil,$$

but then ω was required to choose the β_h so that

$$\frac{r_{h+1}}{r_h} \leq \frac{h(\beta_h)}{h(\beta_{h+1})} + \frac{h(\beta_h)}{r_1 h(\beta_1)} \leq \frac{\omega}{2} + \frac{\omega}{2}.$$

It can now be shown that we have satisfied the conditions necessary to apply the large index proposition and Roth's lemma with $\eta = \omega^{2^{-m+1}} = \varepsilon/4$. This results in

$$m\varepsilon \leq \mathrm{Ind}_{(\boldsymbol{\beta},\boldsymbol{r})}(P) \leq 2m\eta \leq \frac{m\varepsilon}{2}.$$

This is a contradiction.

References