PROOF OF THUE'S THEOREM – PART II

1. Polynomials that vanish to high order at a rational point

Suppose that $P \in \mathbb{Z}[x_1, x_2]$ has the special form

$$P(x_1, x_2) = P_1(x_1)x_2 + P_0(x_1).$$

Suppose that $r \in \mathbb{Q}^2$. If P vanishes to high order at a complicated point r, how big do the coefficients of P have to be? More precisely, we suppose that $\partial_1^j P(r) = 0$ for $0 \le j \le l-1$. Last time we gave two examples. The polynomial $q_2 x_2 - p_2$ which has size $\|r_2\|$, and the polynomial $(q_1 x_1 - p_1)^l$, which has size $\|r_1\|^l$.

By parameter counting it is possible to do somewhat better.

Proposition 1.1. *For any $r \in \mathbb{Q}^2$, and any $l \ge 0$, there is a polynomial $P \in \mathbb{Z}[x_1, x_2]$ with the form $P(x_1, x_2) = P_1(x_1)x_2 + P_0(x_1)$ obeying the following conditions.*

- *$\partial_1^j P(r) = 0$ for $j = 0, ..., l-1$.*
- *$|P| \le C(\epsilon)^l \|r_1\|^{\frac{l}{2}+\epsilon}$, for any $\epsilon > 0$.*
- *The degree of P is $\lesssim \epsilon^{-1}\left(l + \log_{\|r_1\|}\|r_2\|\right)$.*

Proof. We will find our solution by counting parameters. We will choose a degree D, and let P_0, P_1 be polynomials of degree $\le D$. The coefficients of P_0 and P_1 are $\ge 2D$ integer variables at our disposal. We wish to satisfy the l equations

$$\partial_1^j P(r) = 0, j = 0, ..., l-1. \tag{1}$$

After a minor rewriting, each of these equations is a linear equation in the coefficients of P with integer coefficients. If we write $P_1(x_1) = \sum_i b_i x_1^i$ and $P_0(x_1) = \sum_i a_i x^i$, then

$$0 = q_1^D q_2 (1/j!) \partial_1^j P(r) = q_2 (\sum_i b_i \binom{i}{j} p_1^{i-j} q_1^{D-i+j}) + (\sum_i a_i \binom{i}{j} p_1^{i-j} q_1^{D-i+j} p_2).$$

The size of the coefficients in the equations is $\le 2^D \|r_1\|^D \|r_2\|$.

By Siegel's lemma on integer solutions of linear integer equations (in the last lecture), we find a non-zero integer solution of these equations with

$$|P| \le \left[3D \cdot 2^D \|r_1\|^D \|r_2\|\right]^{\frac{l}{2D-l}} \le C^l \|r_1\|^{l\frac{D}{2D-l}} \|r_2\|^{\frac{l}{2D-l}}.$$

We choose $D = 1000\epsilon^{-1}l + 1000\epsilon^{-1}\log_{\|r_1\|}\|r_2\|$. With this value of D, $\frac{D}{2D-l} \le \epsilon/10$, and so the exponent of $\|r_1\|$ is almost $l/2$. Also, the term $\|r_2\|^{\frac{l}{2D-l}} \le \|r_1\|^{\epsilon/10}$. $\square$

Combining our parameter counting with the elementary example $q_2 x_2 - p_2$, we can find P vanishing to order l at r with $|P|$ on the order of $\min(\|r_1\|^{l/2}, \|r_2\|)$. The following result shows that these examples are quite sharp. I believe it is a special case of a lemma of Schneider.

Proposition 1.2. *(Schneider) If $P(x_1, x_2) = P_1(x_1)x_2 + P_0(x_1) \in \mathbb{Z}[x_1, x_2]$, and $r \in \mathbb{Q}^2$, and $\partial_1^j P(r) = 0$ for $j = 0, ..., l-1$, and if $l \ge 2$, then*

$$|P| \ge \min((2DegP)^{-1}\|r_1\|^{\frac{l-1}{2}}, \|r_2\|).$$

Remark. We need to assume that $l \ge 2$ to get any estimate. If we have vanishing only to order 1, then we could have $P(x_1, x_2) = 2x_1 - x_2$, which vanishes at $(r_1, 2r_1)$ for any rational number r_1. As soon as $l \ge 2$, the size of $|P|$ constrains the complexity of r. It can still happen that one component of r is very complicated, but they can't both be very complicated.

Proof. Our assumption is that

$$\partial^j P_1(r_1) r_2 + \partial^j P_0(r_1) = 0, 0 \le j \le l-1.$$

Let $V(x)$ be the vector $(P_1(x), P_0(x))$. Our assumption is that for $0 \le j \le l-1$, the derivatives $\partial^j V(r_1)$ all lie on the line $V \cdot (r_2, 1) = 0$. In particular, any two of these derivatives are linearly dependent. This tells us that many determinants vanish. If V and W are two vectors in $\mathbb{R}^2$, we write $[V, W]$ for the 2×2 matrix with first column V and second column W. Therefore,

$$det[\partial^{j_1} V, \partial^{j_2} V](r_1) = 0, \text{ for any } 0 \le j_1, j_2 \le l-1.$$

Now it follows by the Liebniz rule that

$$\partial_j det[V, \partial V](r_1) = 0, \text{ for any } 0 \le j \le l-2.$$

Remark: Because the determinant is multilinear, we have the Leibniz rule $\partial det[V, W] = det[\partial V, W] + det[V, \partial W]$, which holds for any vector-valued functions $V, W : \mathbb{R} \to \mathbb{R}^2$.

Now $det[V, \partial V]$ is a polynomial in one variable with integer coefficients. If this polynomial is non-zero, then by Gauss's lemma (see last lecture) we conclude that

$$|det[V, \partial V]| \ge \|r_1\|^{l-1}.$$

Expanding out in terms of P, we have $|det[V, \partial V]| = |\partial P_0 P_1 - \partial P_1 P_0| \le 2(DegP)^2 |P|^2$. Therefore, we have $|P| \ge (2DegP)^{-1}\|r_1\|^{\frac{l-1}{2}}$.

The polynomial $det[V, \partial V]$ may also be identically zero. This is a degenerate case, and the polynomial must simplify dramatically. One possibility is that P_1 is identically zero. In this case $P(x_1, x_2) = P_0(x_1)$, and by the Gauss lemma we have that $|P| \geq \|r_1\|^l$. If P_1 is not identically zero, then the derivative of the ratio P_0/P_1 is identically zero. (The numerator of this derivative is $det[V, \partial V]$.) In this case, the polynomial P factors as $(q_2 x_2 - p_2)\tilde{P}(x_1)$, where $\tilde{P}(x_1)$ has integer coefficients. (compare proof of Gauss lemma) In this case, $|P| \geq \|r_2\|$. $\square$

The lower bounds on $|P|$ in this lemma are pretty close to the upper bounds on $|P|$ in the examples above. Speaking informally, both bounds are pretty close to $\min(\|r_1\|^{l/2}, \|r_2\|)$.

2. Polynomials that vanish at algebraic points

Our whole discussion can be generalized in a straightforward way to algebraic points instead of rational points. In the proof of Thue's theorem, we have an algebraic number β, and r_1 and r_2 are rational numbers that approximate β with very large heights. The point (r_1, r_2) is close to (β, β). We are going to compare finding an integral polynomial that vanishes to high order at (β, β) and finding an integral polynomial that vanishes to high order at (r_1, r_2).

By using parameter counting, we will see that there is an integral polynomial vanishing to high order at (β, β) whose coefficients are much smaller than what we could find for a polynomial vanishing to high order at (r_1, r_2).

Proposition 2.1. *Let $\beta \in \mathbb{R}$ be an algebraic number. For any natural number l, and any $\epsilon > 0$, there is a polynomial $P \in \mathbb{Z}[x_1, x_2]$ with the form $P(x_1, x_2) = P_1(x_1)x_2 + P_0(x_1)$ with the following properties.*

- $\partial_1^j P(\beta, \beta) = 0$ *for* $0 \leq j \leq l - 1$.
- $|P| \leq C(\beta)^{l/\epsilon}$.
- *The degree of P is* $\leq (1 + \epsilon)(1/2)deg(\beta)l + 1$.

Proof. This Proposition follows by the same parameter counting argument as above. There is one significant new idea in order to deal with algebraic numbers. We let D a degree to choose later. As above, we write $P_1(x) = \sum_{i=0}^{D} b_i x^i$ and $P_0(x) = \sum_{i=0}^{D} a_i x^i$. The coefficents a_i and b_i are $\geq 2D$ integer variables at our disposal. For each $0 \leq j \leq l - 1$, our vanishing equation is

$$0 = (1/j!)\partial_1^j P(\beta, \beta) = \sum_i b_i \binom{i}{j} \beta^{i-j+1} + \sum_i a_i \binom{i}{j} \beta^{i-j}. \quad (1)$$

This is a linear equation in a_i, b_i with coefficients in $\mathbb{Z}[\beta]$. We will see that it is equivalent to $deg(\beta)$ linear equations with coefficients in $\mathbb{Z}$. Since β is an algebraic number, we will check that $1, \beta, ..., \beta^{deg(\beta)-1}$ form a basis for the vector space $\mathbb{Q}[\beta]$ over the field $\mathbb{Q}$. In particular, any power β^i can be expanded as a rational combination of $1, \beta, ..., \beta^{deg(\beta)-1}$. Substituting in, we can rewrite equation (1) in the form:

$$0 = \sum_{k=0}^{deg(\beta)-1} \beta^k \left[\sum_i b_i B_{ik} + \sum_i a_i A_{ik} \right] = 0,$$

where A_{ik} and B_{ik} are rational numbers. Since $1, \beta, ..., \beta^{deg(\beta)-1}$ are linearly independent over $\mathbb{Q}$, this list of equations is equivalent to the $deg(\beta)$ equations

$$\sum_i b_i B_{ik} + \sum_i a_i A_{ik} = 0, \text{ for all } 0 \le k \le deg(\beta) - 1. \tag{2}$$

After multiplying by a large constant to clear the denominators, we get $deg(\beta)$ equations with integer coefficients. In total, our original l equations $\partial_1^j P(r) = 0$ for $j = 0, ..., l-1$ are equivalent to $deg(\beta)l$ integer linear equations in the coefficients of P. Since we have $> 2D$ coefficients, we can find a non-trivial integer solution as long as $D \ge (1/2)deg(\beta)l$.

Our next task is to estimate the size of the solution. To do this, we need to estimate the heights of the coefficients A_{ik}, B_{ik}. Also we get a much better estimate by taking D slightly larger than $(1/2)deg(\beta)l$, and for this reason we choose D to be the least integer $\ge (1+\epsilon)(1/2)deg(\beta)l$. To estimate the heights of A_{ik}, B_{ik}, we consider more carefully how to expand β^d in terms of $1, \beta, ..., \beta^{d-1}$.

Lemma 2.2. *Suppose $Q(\beta) = 0$, where $Q \in \mathbb{Z}[x]$ with degree $deg(Q) = deg(\beta)$ and leading coefficient $q_{deg(beta)}$. Then for any $d \ge 0$, we can write*

$$q_{deg(\beta)}^d \beta^d = \sum_{k=0}^{deg(\beta)-1} A_{kd} \beta^k,$$

where $A_{kd} \in \mathbb{Z}$ and $|A_{kd}| \le [2|Q|]^d$.

Proof. We have $0 = Q(\beta) = \sum_{k=0}^{deg(\beta)} q_k \beta^k$. We do the proof by induction on d, starting with $d = deg(\beta)$. For $d = deg(\beta)$, the equation $Q(\beta) = 0$ directly gives

$$q_{deg(beta)}^{deg(\beta)} \beta^{deg(\beta)} = \sum_{k=0}^{deg(\beta)-1} (-q_k) \beta^k. \tag{$*$}$$

If we multiply both sides by $q_{deg(\beta)}^{deg(\beta)-1}$, we get a good expansion for the case $d = deg(\beta)$. Now we proceed by induction. Suppose that $q_{deg(\beta)}^d \beta^d = \sum_{k=0}^{deg(\beta)-1} A_{kd} \beta^k$. Multiplying by $q_{deg(\beta)} \beta$, we get

$$q_{deg(\beta)}^{deg(\beta)+1}\beta^{deg(\beta)+1} = \sum_{k=0}^{deg(\beta)-1} A_{kd} q_{deg(\beta)} \beta^{k+1} = \sum_{k=1}^{deg(\beta)-1} A_{k-1,d} q_{deg(\beta)} \beta^k + \sum_{k=0}^{deg(\beta)-1} A_{deg(\beta)-1,d}(-q_k)\beta^k.$$

□

Plugging in this lemma, we see that $q_{deg(\beta)}^D A_{ik}, q_{deg(\beta)}^D B_{ik}$ are integers of size $\leq D[2|Q|]^D$. The integer matrix that we are solving has coefficients of size $\leq D[2|Q|]^D$. It is a matrix with dimensions $(2D+2) \times deg(\beta)l$, and so it has operator norm $\leq (2D+2)D[2|Q|]^D \leq C(\beta)^D$.

Now applying Siegel's lemma, we see that we can find an integer solution P with $|P|$ bounded by

$$C(\beta)^{D\frac{deg(\beta)l}{2D-deg(\beta)l}} \leq C(\beta)^{D/\epsilon}.$$

Since $D \leq C(\beta)l$, we can redefine $C(\beta)$ so that $|P| \leq C(\beta)^{l/\epsilon}$.

□

3. Summary

Suppose that β is an algebraic number, and that r_1, r_2 are two very good rational approximations of β. We may suppose that $\|r_1\|$ is very large and $\|r_2\|$ is (much) larger. Say $\|r_2\| \sim \|r_1\|^m$.

We consider polynomials $P \in \mathbb{Z}[x_1, x_2]$ of the simple form $P(x_1, x_2) = P_1(x_1)x_2 + P_0(x_1)$. We can arrange that $\partial_1^j P(\beta, \beta) = 0$ for $0 \leq j \leq m-1$ with $|P| \leq C(\beta)^m$. On the other hand, if $\partial_1^j P(r) = 0$ for $0 \leq j \leq l-1$, then we must have $|P| \gtrsim \|r_1\|^{l/2}$. Since $\|r_1\|$ is much larger than $C(\beta)$, it follows that l must be much smaller than m. This creates a certain tension.

As we will see, if r was too close to (β, β), than P would have to vanish too much at r, giving a contradiction.