**Roth's Theorem**

## 0.1 The Proof of Roth' Theorem

**Theorem (Roth)** *Let $\alpha$ be an algebraic number of degree $\geq 2$. Then, for every $\epsilon > 0$, the inequality*

$$\left|\frac{p}{q} - \alpha\right| > \frac{1}{q^{2+\epsilon}}$$

*holds for all, except for finitely many, rational numbers $p/q$.*

To prove Roth's theorem, we first state several lemmas. The first one is the so-called Siegel's lemma. Siegel's lemma is a corollary of the "pigeonhole principle."

**Lemma 1(Siegel's Lemma)** *Let $A$ be an $M \times N$ matrix with $M < N$ and having entries in $\mathbf{Z}$ of absolute value at most $Q$, where $\mathbf{Z}$ is the set of integers. Then there exists a nonzero vector $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbf{Z}^N$ with $A\mathbf{x} = 0$, such that*

$$|x_i| \leq [(NQ)^{M/(N-M)}] =: Z, \qquad i = 1, \ldots, N.$$

*Proof* The number of integer points in the box

$$0 \leq x_i \leq Z, \quad i = 1, \ldots, N$$

is $(Z+1)^N$. On the other hand, for all $j = 1, \ldots, N$ and for each such $\mathbf{x}$, the $j^{th}$ coordinate $y_j$ of the vector $\mathbf{y} := A\mathbf{x}$ lies in the interval $[-n_jQZ, (N - n_j)QZ]$, where $n_j$ is the number of negative entries in the $j^{th}$ row of $A$. Therefore, there are at most $(NQZ+1)^M < (Z+1)^N$ possible values of $A\mathbf{x}$. Hence, there must exist vectors $\mathbf{x}_1 \neq \mathbf{x}_2$ in the box and such that $A\mathbf{x}_1 = A\mathbf{x}_2$. Then $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ satisfies the conditions of the lemma.

The second lemma states that most values $i_1/d_1 + \cdots + i_n/d_n$ with $0 \leq i_h \leq d_h$ $(h = 1, \ldots, m$ are close to $n/2$.

**Lemma 2 (A Combinatorial Lemma)** Let $d_1, \ldots, d_n$ be integers greater than or equal to 1 and let $\epsilon_1 > 0$. The number of sets of integers $(i_1, \ldots, i_n)$ satisfying

$$0 \leq i_1 \leq d_1, \ldots, 0 \leq i_n \leq d_n$$

and

$$\left| \sum_{h=1}^{n} \frac{i_h}{d_h} - \frac{n}{2} \right| \geq \epsilon n$$

is at most $(d_1 + 1) \ldots (d_n + 1)/(4n\epsilon^2)$.

*Proof.* We may consider $i_1, \ldots, i_n$ as independent stochastic variables such that $i_h$ is uniformly distributed on $\{0, \ldots, d_h\}$. Define the stochastic variable $X = \sum_{h=1}^{n} i_h/d_h$. Then $X$ has expectation $\mu = n/2$ and variance

$$\sigma^2 = Var(i_1/d_1) + \cdots + Var(i_n/d_n).$$

We have

$$Var(i_h/d_h) = \sum_{i_h=0}^{d_h} \left( \frac{i_h}{d_h} - \frac{1}{2} \right)^2 \frac{1}{d_h + 1} = \frac{2d_h + 1}{6d_h} - \frac{1}{4} \leq \frac{1}{4}.$$

Hence $\sigma^2 \leq n/4$. By Kolmogorov's generalization of Chebyshev's inequality, we have $\mathrm{Prob}(|X - \mu| \geq c) \leq \sigma^2/c^2$. Thus

$$Prob(|X - n/2| \geq \epsilon_1 m) \leq \frac{1}{4m\epsilon_1^2}.$$

This proves the lemma

**Definition 1** *For a polynomial* $P(X_1, \ldots, X_n) \in \mathbf{Z}[X_1, \ldots, X_n]$ *and* $\mathbf{i} = (i_1, \ldots, i_n) \in \mathbf{Z}_{\geq 0}^n$*, put*

$$P_{\mathbf{i}}(X_1, \ldots, X_n) = \frac{1}{i_1! \cdots i_n!} \frac{\partial^{i_1}}{\partial X_1^{i_1}} \cdots \frac{\partial^{i_n}}{\partial X_1^{i_n}} P(X)$$

$$= \sum_{l_1, \ldots, l_n \geq 0} \begin{pmatrix} j_1 \\ i_1 \end{pmatrix} \cdots \begin{pmatrix} j_n \\ i_n \end{pmatrix} C(j_1, \ldots, j_n) X_1^{j_1 - i_1} \cdots X_n^{j_n - i_n}.$$

*Let $\alpha_1, \ldots, \alpha_n \in \mathbf{C}$ and let $d_1, \ldots, d_n$ be positive integers. Then the* **index of** $P$ **at** $\alpha = (\alpha_1, \ldots, \alpha_n)$ **with weights** $d_1, \ldots, d_n$ *is*

$$t(P, (\alpha_1, \ldots, \alpha_m), d_1, \ldots, d_n) = \min \left\{ \sum_{i=1}^{n} \frac{l_i}{d_i} \;\middle|\; P_{\mathbf{i}}(\alpha) \neq 0 \right\}.$$

Note that $i(PQ) = i(P) + i(Q)$ and $i(P+Q) \geq \min\{i(P), i(Q)\}$.

The third lemma provides the construction of a polynomial with high index at some given point. For $P \in \mathbf{Z}[X_1, \ldots, X_n]$ we denote the maximum of the absolute value of the coefficients of $P$ by $|P|$.

**Lemma 3(the Index Theorem)** *Suppose that $\alpha$ is an algebraic integer of degree $d \geq 2$. Let $\epsilon > 0$, and let $n$ be an integer with $n \geq d/2\epsilon^2$. Let $d_1, \ldots, d_n$ be positive integers. Then there is $P \in \mathbf{Z}[X_1, \ldots, X_n]$, $P \not\equiv 0$, such that*
*(i) $P$ has degree $\leq d_h$ in $X_h$,*
*(ii)*

$$t(P, (\alpha, \ldots, \alpha), d_1, \ldots, d_n) \geq n(1-\epsilon)/2$$

*(iii) $|P| \leq C_1^{d_1+\cdots+d_n}$.*

*Proof.* Write $P(X_1, \ldots, X_n) = \sum_{j_1=0}^{d_1} \cdots \sum_{j_n=0}^{d_n} z(j_1, \ldots, j_n) X_1^{j_1} \cdots X_n^{j_n}$, where $z(j_1, \ldots, j_n)$ are the integers which have to be determined such that $(ii)$ hodls, i.e. $P_{\mathbf{i}}(\alpha) = 0$ for $i_1/d_1 + \cdots i_n/d_n \leq n(1-\epsilon)/2$. By taking all these expression together, we obtain

$$A_0\mathbf{z} + \alpha A_1 \mathbf{z} + \cdots + \alpha^{d_1+\cdots+d_n} A_{d_1+\cdots+d_n} \mathbf{z} = \mathbf{0}$$

where $A_i$ are $M \times N$ integer matrices with $|A_i| \leq 4^{d_1+\cdots+d_n}$, where $N = (d_1 +!) \cdots (d_m + 1)$ and $M$ is the number of tuples $\mathbf{i}$ with $i_1/d_1 + \cdots i_n/d_n \leq n(1-\epsilon)/2$. Using the fact that $\alpha$ is an algebra number of degree $d$, we get

$$B_0\mathbf{z} + \alpha B_1\mathbf{z} + \cdots + \alpha^{d-1} B_{d-1}\mathbf{z} = \mathbf{0}$$

where $B_i$ are $M \times N$ integer matrices with $|B_i| \leq C_2^{d_1+\cdots+d_n}$. Since $1, \alpha, \alpha^2, \ldots, \alpha^{d-1}$ are $\mathbf{Z}$-linear independent, we have $B_0\mathbf{z} = 0, \ldots, B_{d-1}\mathbf{z} = \mathbf{0}$. Hence $B\mathbf{z} = \mathbf{0}$ where $B$ is an $dM \times N$ integer matrices with $|B| \leq C_2^{d_1+\cdots+d_n}$. By the combinational lemma, we have

$$M \leq \frac{(d_1+1)\cdots(d_n+1)}{4m\epsilon^2} = \frac{N}{4m\epsilon^2} \leq \frac{N}{2d}.$$

Now Siegel's lemma implies that there is a non-zero integer vector $\mathbf{z}$ such that $B\mathbf{z} = \mathbf{0}$ and

$$|\mathbf{z}| \leq (N|B|)^{dM/(N-dM)} \leq N|B| \leq C_3^{d_1+\cdots+d_n}.$$

Note that the constants $C_1, C_2$ and $C_3$ depend only on $\alpha$. This finishes the proof.

The fourth lemma gives a sufficient condition for a polynomial to have small index with respect to the approximation vector $(p_1/q_1, \ldots . p_n/q_n)$ and $(d_1, \ldots, d_n)$.

**Lemma 4 (Roth's Lemma)** *Let $n \geq 1$ be a positive integer, and $\epsilon > 0$. There exists a number $C_4 = C_4(m, \epsilon) > 1$ with the following property: Let $d_j (j = 1, \ldots, n)$ be integers with $d_h \geq C_4 d_{h+1}, h = 1, \ldots, n-1$. Let $(p_1, q_1), \ldots, (p_n, q_n)$ be pairs of coprime integers with $q_h^{d_h} \geq q_1^{d_1}$ and $q_h \geq 2^{2mC_4}, h = 1, \ldots, n$. Let $P(X_1, \ldots, X_n) \not\equiv 0$ be a polynomial in $\mathbf{Z}[X_1, \ldots, X_n]$ of degree at most $d_h$ in $X_h$ with*

$$|P|^{C_4} \leq q_1^{d_1}, \quad P \not\equiv 0.$$

*Then*

$$t = t(P, (p_1/q_1, \ldots, p_n/q_n), d_1, \ldots, d_n) \leq \epsilon.$$

The proof of Roth's lemma can be found in Lang's book. We omit it here.

**Proof of Roth's Theorem** Assume that Roth's Theorem fails, i.e.

$$\left| \frac{p}{q} - \alpha \right| < \frac{1}{q^{2+\delta/2}} \qquad (*)$$

holds for infinitely many $p/q$. We will derive a contradiction. Without loss of generality, we assume that $\alpha$ is an algebraic integer of degree $d \geq 2$ with $|\alpha| < 1$. We also assume that $0 < \delta < 1/2$.

**Step 1**: Choice of "suitable" points $p_h/q_h$. Let $P$ be the polynomial constructed in the index Theorem with respect to $\alpha$, $\epsilon = \delta/12$, $n > d/2\epsilon^2$ and arbitrary $d_1, \ldots, d_n$. Then $P$ has index $> m(1-\epsilon)/2$ with respect to $(\alpha, \ldots, \alpha)$ and $(d_1, \ldots, d_n)$. We first chose solutions $p_1/q_1, \ldots, p_n/q_n$ as follows:

(a) Choose $(p_1, q_1)$ with

$$q_1 > \max((6C_1)^{1/\epsilon}, C_1^m, 2^{2mC_4})$$

where $C_1$ is the constant appearing in the index Theorem, and $C_4$ is the constant appearing in the Roth's lemma.

(b) Choose solutions $(p_2, q_2), \ldots, (p_n, q_n)$ such that

$$q_{h+1} > q_h^{(1+\epsilon)C_4}, \quad 1 \leq h \leq n-1$$

(c) Choose $d_1$ so that

$$q_1^{\epsilon d_1} \geq q_n$$

(d) for $q = 2, \ldots, n$, choose $d_h$ such that

$$q_1^{d_1} \leq q_h^{d_h} < q_1^{d_1(1+\epsilon)}.$$

(This is possible since $q_1^{\epsilon d_1} \geq q_n \geq q_h$).

It is easy to verify that Roth's lemma are satisfied using the above choice.

**Step 2**: We show, using the Taylor's expansion, that $P(p_1/q_1, \ldots, p_n/q_n) = 0$. In fact, we can show a stronger result that $P$ has index $> \epsilon$ with respect to $(p_1/q_1, \ldots, p_n/q_n)$ and $(d_1, \ldots, d_n)$. To do so, we need to prove that for $\mathbf{i}$ with

$$\frac{i_1}{d_1} + \cdots + \frac{i_n}{d_n} \leq \epsilon$$

we have $P_{\mathbf{i}}(p_1/q_1, \ldots, p_n/q_n) = 0$. Note that

$$P_{\mathbf{i}}(\alpha) = \sum_{\mathbf{j}} P_{\mathbf{j}}(\mathbf{0}) \left( \begin{array}{c} j_1 \\ i_1 \end{array} \right) \cdots \left( \begin{array}{c} j_n \\ i_n \end{array} \right) \alpha^{j_1 - i_1} \cdots \alpha^{j_n - i_n},$$

whence, using $|\alpha| < 1$,

$$|P_{\mathbf{i}}(\alpha)| \leq |P| \max_{\mathbf{j}} \left( \begin{array}{c} j_1 \\ i_1 \end{array} \right) \cdots \left( \begin{array}{c} j_n \\ i_n \end{array} \right) \leq (2C_1)^{d_1 \cdots + d_n} \leq (2C_1)^{nd_1}$$

where the maximum extends over all $\mathbf{j}$ with $j_h \leq d_h$ for $h = 1, \ldots, n$. Expand $P_{\mathbf{i}}(X)$ in a Taylor series around $(\alpha, \ldots, \alpha)$,

$$P_{\mathbf{i}}(X) = \sum_{\mathbf{j}} P_{\mathbf{j}}(\alpha) \left( \begin{array}{c} j_1 \\ i_1 \end{array} \right) \cdots \left( \begin{array}{c} j_n \\ i_n \end{array} \right) (X_1 - \alpha)^{j_1 - i_1} \cdots (X_n - \alpha)^{j_n - i_n}.$$

By the construction of $P$ (see condition (ii) in the index theorem),

$$t(P, (\alpha, \dots, \alpha), d_1, \dots, d_n) \geq n(1-\epsilon)/2.$$

Hence $P_{\mathbf{j}}(\alpha, \dots, \alpha) = 0$, if $j_1/d_1 + \dots j_n/d_n \leq n(1-\epsilon)/2$, so certainly if $(j_1 - i_1)/d_1 + \dots + (j_n - i_n)/d_n \leq n(1-3\epsilon)/2$. Furthermore,

$$\sum_{\mathbf{j}} P_{\mathbf{j}}\alpha) \begin{pmatrix} j_1 \\ i_1 \end{pmatrix} \cdots \begin{pmatrix} j_n \\ i_n \end{pmatrix} \leq (2C_1)^{nd_1} \sum_{\mathbf{j}} 2^{j_1+\dots+j_n} \leq (6C_1)^{nd_1}.$$

Hence, for

$$F(X) := P_{\mathbf{i}}(X) = \sum_{(l)\geq 0} F^{(l)}(\alpha, \dots, \alpha)(X-\alpha)^{(l)}.$$

we have that all the terms will be 0 except those belonging to $(l)$ with

$$\frac{l_1}{d_1} + \dots + \frac{l_n}{d_n} \geq n(1-3\epsilon)/2$$

and

$$\sum_{\mathbf{j}} |F^{(l)}(\alpha, \dots, \alpha)| \leq (6C_1)^{nd_1}.$$

It follows that, on denoting by (*) with the $(l)$ with $\frac{l_1}{d_1} + \dots + \frac{l_n}{d_n} \geq n((1-3\epsilon)/2$,

$$\begin{aligned}
\log |F(p_1/q_1, \dots, p_n/q_n)| &\leq (6C_1)^{nd_1} max^*_{(l)} \left|\frac{p_1}{q_1} - \alpha\right|^{l_1} \cdots \left|\frac{p_n}{q_n} - \alpha\right|^{l_n} \\
\leq & \quad (6C_1)^{nd_1} max^*_{(l)}((q_1^{d_1})^{l_1/d_1} \cdots (q_n^{d_n})^{l_n/d_n})^{-2-\delta} \\
\leq & \quad (6C_1)^{nd_1} max^*_{(l)}((q_1^{d_1})^{(l_1/d_1 + \dots + q_n^{d_n})(-2-\delta)} \\
\leq & \quad (q_1)^{\epsilon nd_1}(q_1^{d_1})^{-n(1-3\epsilon)(1+\delta/2)} \\
\leq & \quad (q_1^{d_1} \cdots q_n^{d_n})^{\{\epsilon - -n(1-3\epsilon)(1+\delta/2)\}/(1+\epsilon)} \\
< & \quad (q_1^{d_1} \cdots q_n^{d_n})^{-1}.
\end{aligned}$$

On the other hand, $|F(p_1/q_1, \dots, p_n/q_n)|$ is a rational number with denominator dividing $q_1^{d_1} \cdots q_n^{d_n}$. Thus

$$P_{\mathbf{i}}(p_1/q_1, \dots, p_n/q_n) = F(p_1/q_1, \dots, p_n/q_n) = 0.$$

**Step 3**: The conclusion in Step 2 contradicts with the Roth's lemma. So this proves Roth's theorem.