1. Outline of the proof of Thue’s theorem

**Theorem 1.1. (Thue)** If \( \beta \) is an irrational algebraic number, and \( \gamma > \frac{\deg(\beta)+2}{2} \), then there are only finitely many integer solutions to the inequality

\[
|\beta - \frac{p}{q}| \leq |q|^{-\gamma}.
\]

By using parameter counting, we constructed polynomials \( P \) with integer coefficients that vanish to high order at \((\beta, \beta)\). The degree of \( P \) and the size of \( P \) are controlled.

If \( r_1, r_2 \) are rational numbers with large height, then we proved that \( P \) cannot vanish to such a high order at \( r = (r_1, r_2) \). For some \( j \) of controlled size, we have \( \partial_j^P(r) \neq 0 \). Since \( P \) has integer coefficients, and \( r \) is rational, \( |\partial_j^P(r)| \) is bounded below.

Since \( P \) vanishes to high order at \((\beta, \beta)\), we can use Taylor’s theorem to bound \( |\partial_j^P(r)| \) from above in terms of \(|\beta - r_1|\) and \(|\beta - r_2|\). So we see that \(|\beta - r_1|\) or \(|\beta - r_2|\) needs to be large.

Here is the framework of the proof. We suppose that there are infinitely many rational solutions to the inequality \(|\beta - r| \leq \|r\|^{-\gamma} \). Let \( \epsilon > 0 \) be a small parameter we will play with. We let \( r_1 \) be a solution with very large height, and we let \( r_2 \) be a solution with much larger height. Using these, we will prove that \( \gamma \leq \frac{\deg(\beta)+2}{2} + C(\beta)\epsilon \).

2. The polynomials

For each integer \( m \geq 1 \), we proved that there exists a polynomial \( P = P_m \in \mathbb{Z}[x_1, x_2] \) with the following properties:

1. We have \( \partial_j^P(\beta, \beta) = 0 \) for \( j = 0, ..., m - 1 \).
2. We have \( \deg_2 P \leq 1 \) and \( \deg_1 P \leq (1 + \epsilon)\frac{\deg(\beta)}{2}m \).
3. We have \( |P| \leq C(\beta, \epsilon)^m \).

3. The rational point

Suppose that \( r_1, r_2 \) are good rational approximations to \( \beta \) in the sense that

\[
\|\beta - r_i\| \leq \|r_1\|^{-\gamma}.
\]
Also, we will suppose that \( \| r_1 \| \) is sufficiently large in terms of \( \beta, \epsilon \), and that \( \| r_2 \| \) is sufficiently large in terms of \( \beta, \epsilon \), and \( \| r_1 \| \).

If \( l \geq 2 \) and \( \partial_1^j P(r) = 0 \) for \( j = 0, \ldots, l - 1 \), then we proved the following estimate:

\[
|P| \geq \min((2 \deg P)^{-1} \| r_1 \|^{\frac{l-1}{2}}, \| r_2 \|).
\]

Given our bound for \( |P| \), we get

\[
C(\beta, \epsilon)^m \geq \min(\| r_1 \|^{\frac{l-1}{2}}, \| r_2 \|).
\]

From now on, we only work with \( m \) small enough so that

\[
C(\beta, \epsilon)^m < \| r_2 \|.
\]

Therefore, \( \| r_1 \|^{\frac{l-1}{2}} \leq C(\beta, \epsilon)^m \). We assume that \( \| r_1 \| \) is large enough so that \( \| r_1 \|^\epsilon > C(\beta, \epsilon) \), and this implies that \( l \leq \epsilon m \). Therefore, there exists some \( j \leq \epsilon m \) so that \( \partial_1^j P(r) \neq 0 \).

Let \( \tilde{P} = (1/j!)(\partial_1^j P) \). The polynomial \( \tilde{P} \) has integer coefficients, and \( |\tilde{P}| \leq 2^{\deg P} |P| \).

Therefore, \( \tilde{P} \) obeys essentially all the good properties of \( P \) above:

1. We have \( \partial_1^j \tilde{P}(\beta, \beta) = 0 \) for \( j = 0, \ldots, (1-\epsilon)\epsilon m - 1 \).
2. We have \( \deg_2 \tilde{P} \leq 1 \) and \( \deg_1 \tilde{P} \leq (1+\epsilon)\frac{\deg(\beta)}{2} m \).
3. We have \( |\tilde{P}| \leq C(\beta, \epsilon)^m \).
4. We also have \( \tilde{P}(r) \neq 0 \).

Since \( \tilde{P} \) has integer coefficients, we can write \( \tilde{P}(r) \) as a fraction with a known denominator: \( q_1^{\deg_1 \tilde{P}} q_2^{\deg_2 \tilde{P}} \). Therefore,

\[
|\tilde{P}(r)| \geq \| r_1 \|^{-\deg_1 \tilde{P}} \| r_2 \|^{-\deg_2 \tilde{P}} \geq \| r_1 \|^{-\deg_1 \tilde{P} - (1+\epsilon)\frac{\deg(\beta)}{2} m \| r_2 \|^{-1}}.
\]

We make some notation to help us focus on what’s important. In our problem, terms like \( \| r_1 \|^m \) or \( \| r_2 \| \) are substantial, but terms like \( \| r_1 \|^{\epsilon m} \) or \( \| r_1 \| \) are minor in comparison. Therefore, we write \( A \lesssim B \) to mean \( A \leq \| r_1 \|^{am} \| r_1 \|^{b} \), for some constants \( a, b \) depending only on \( \beta \).

Recall that \( \| r_1 \|^\epsilon \) is bigger than \( C(\beta, \epsilon) \), so \( C(\beta, \epsilon)^m \lesssim 1 \). Our main inequality for this section is

\[
|\tilde{P}(r)| \gtrsim \| r_1 \|^{-\frac{\deg(\beta)}{2} m \| r_2 \|^{-1}}.
\]

4. Taylor’s theorem estimates

We recall Taylor’s theorem.
**Theorem 4.1.** If \( f \) is a smooth function on an interval, then \( f(x + h) \) can be approximated by its Taylor expansion around \( x \):
\[
f(x + h) = \sum_{j=0}^{m-1} \frac{1}{j!} \partial_j f(x) h^j + E,
\]
where the error term \( E \) is bounded by
\[
|E| \leq \left( \frac{1}{m!} \right) \sup_{y \in [x,x+h]} |\partial_m f(y)|.
\]

In particular, if \( f \) vanishes to high order at \( x \), then \( f(x + h) \) will be very close to \( f(x) \).

**Corollary 4.2.** If \( Q \) is a polynomial, and \( Q \) vanishes at \( x \) to order \( m \geq 1 \), and if \( |h| \leq 1 \), then
\[
|Q(x + h)| \leq C(x)^{\deg Q} |Q|h^m.
\]

**Proof.** We see that \( (1/m!) \partial^m Q \) is a polynomial with coefficients of size \( \leq 2^{\deg Q} |Q| \).
We evaluate it at a point \( y \) with \( |y| \leq |x| + 1 \). Each monomial has norm \( \leq 2^{\deg Q} |Q|(|x| + 1)^{\deg Q} \), and there are \( \deg Q \) monomials. \( \square \)

Let \( Q(x) = \tilde{P}(x, \beta) \). The polynomial \( Q \) vanishes to high order \((1 - \epsilon)m \) at \( x = \beta \), and \( |Q| \leq C(\beta, \epsilon)^m \).

From the corollary we see that
\[
|\tilde{P}(r_1, \beta)| \leq C(\beta, \epsilon)^m |\beta - r_1|^{(1-\epsilon)m}.
\]

On the other hand, \( \partial_2 \tilde{P} \) is bounded by \( C(\beta, \epsilon)^m \) in a unit disk around \((\beta, \beta)\), and so
\[
|\tilde{P}(r_1, r_2) - \tilde{P}(r_1, \beta)| \leq C(\beta, \epsilon)^m |\beta - r_2|.
\]

Combining these, we see that
\[
|\tilde{P}(r)| \lesssim |\beta - r_1|^{(1-\epsilon)m} + |\beta - r_2| \lesssim \|r_1\|^{-\gamma m} + \|r_2\|^{-\gamma}.
\]

5. **Putting it together**

As long as \( \|r_1\|^\epsilon > C(\beta, \epsilon) \) and \( \|r_2\| > C(\beta, \epsilon)^m \), we have proven the following inequality:
\[
\|r_1\|^{-\deg Q m} \|r_2\|^{-1} \lesssim \|r_1\|^{-\gamma m} + \|r_2\|^{-\gamma}
\]

Now we can choose \( m \). As \( m \) increases, the right-hand side decreases until \( \|r_1\|^m \sim \|r_2\| \), and then the \( \|r_2\|^{-\gamma} \) term becomes dominant. Therefore, we choose \( m \) so that
\[
\|r_1\|^m \leq \|r_2\| \leq \|r_1\|^{m+1}.
\]
We see that \( \|r_2\| \geq \|r_1\|^m > C(\beta, \epsilon)^m \), so the assumption about \( r_2 \) and \( m \) above is satisfied. The inequality becomes
\[
\|r_1\|^{-\frac{\deg(\beta) - \gamma m}{2}} \lesssim \|r_1\|^{-\gamma m}.
\]
Multiplying through to make everything positive, we get
\[
\|r_1\|^{\gamma m} \lesssim \|r_1\|^\frac{\deg(\beta) + 2}{2} m.
\]
Unwinding the \( \lesssim \), this actually means
\[
\|r_1\|^\gamma m \leq \|r_1\|^\frac{b + a \epsilon + \deg(\beta) + 2}{2} m.
\]
(If we had been more explicit, we could have gotten specific values for \( a, b \), but it doesn’t matter much.)

Taking the logarithm to base \( \|r_1\| \) and dividing by \( m \), we get
\[
\gamma \leq \frac{(b/m) + a \epsilon + \deg(\beta) + 2}{2}.
\]

If \( \|r_2\| \) is large enough compared to \( \|r_1\| \), then \( (1/m) \leq \epsilon \), and we have \( \gamma \leq \frac{(a + b)\epsilon + \deg(\beta) + 2}{2} \). Taking \( \epsilon \to 0 \) finishes the proof.