An unexpected application of the Cauchy-Schwarz inequality

Titu Andreescu, Gabriel Dospinescu

The purpose of this expository paper is to present a beautiful application of the Cauchy-Schwarz inequality to number theory. More precisely, following [1] (which is a slightly expanded version of the proof presented in [3]) we will present an elementary and self-contained proof of the following

**Theorem 1.** Let \( a, b > 1 \) be integers such that for all prime powers \( q \) there is \( k \geq 1 \) (depending on \( q \)) such that \( b \equiv a^k \pmod{q} \). Then \( b \) is an integral power of \( a \).

Before delving into the proof of the previous theorem, let us note that the result does not hold if instead of prime powers \( q \) we limit ourselves to primes. Indeed, we have the following classical result, whose proof we give for the convenience of the reader:

**Proposition 2.** For all primes \( p \) one can find an integer \( x \) such that \( x^8 \equiv 16 \pmod{p} \).

In other words, 16 is an eighth power modulo all primes, but it is clearly not the eighth power of an integer.

**Proof.** Note that

\[
x^8 - 16 = (x^4 - 4)(x^4 + 4) = (x^2 - 2)(x^2 + 2)(x^2 - 2x + 2)(x^2 + 2x + 2),
\]

thanks to Sophie Germain’s identity. This can be rewritten as

\[
x^8 - 16 = (x^2 - 2)(x^2 + 2)((x - 1)^2 + 1)((x + 1)^2 + 1).
\]

Hence we need to prove that for all primes \( p \) one of the numbers \(-1, 1, -2, 2\) is a quadratic residue mod \( p \). This follows from the fact that \((-1)(-2)^2 = 2^2\) is a square mod \( p \) and from the multiplicativity of Legendre’s symbol.

Finally, let us mention that using the Chebotarev density theorem, one can prove the following general result (for which we refer to the beautiful papers [4] and [2], or to theorem 9.B.60 in [1]).

**Theorem 3.** Let \( n > 1 \) be an integer and let \( a \) be an integer such that \( a \) is an \( n \)th power modulo any sufficiently large \(^1\) prime. Then either \( a \) is the \( n \)th power of an integer or \( 8 | n \) and \( a = 2^k b^n \) for some integer \( b \).

\(^1\)The proofs show that it suffices to assume that this holds for a set of primes of Dirichlet density 1.
Before going to the heart of the proof of theorem 1, let us make a preliminary reduction:

**Proposition 4.** Suppose that \(a, b\) satisfy the conditions of theorem 1, and that \(\log a\) and \(\log b\) are linearly dependent over \(\mathbb{Q}\). Then \(b\) is an integral power of \(a\).

Let us mention that the condition of linear dependence simply means that \(a^i b^j = 1\) for some rational numbers \(i, j\), not both zero.

**Proof.** Using the unique factorization theorem, it follows from the linear dependence of \(\log a\) and \(\log b\) that we can write \(a = c^i\) and \(b = c^j\) for some integer \(c > 1\) and some relatively prime positive integers \(i, j\). If \(q\) is a prime power dividing \(c^i - 1\), then by hypothesis one can find \(k \geq 1\) (depending on \(q\)) such that \(q\) divides \(c^j - c^k i\). Thus \(q\) divides \(c^j - 1\) and so, again by unique factorization, \(c^j - 1\) divides \(c^i - 1\). A standard argument based on euclidean division shows that this forces \(i | j\) and so \(b\) is an integral power of \(a\). \(\square\)

Thus, proving theorem 1 comes down to proving that \(\log a\) and \(\log b\) are linearly dependent. The proof of this result uses a very clever application of the Cauchy-Schwarz inequality, known as Gallagher’s sieve. Sieves are a very powerful tool in analytic number theory, and there is a huge literature on their refinements and applications. See for instance \([3, 5, 6, 7, 8]\) (for an elementary exposition of basic facts see the addendum to chapter 2 in \([1]\) ). Roughly speaking, sieve theory is a powerful tool serving to answer the following problem: suppose that \(A\) is a set of integers such that \(A \pmod{p} = \{x \pmod{p}|x \in A\}\) is relatively small for all primes \(p\) in a finite set \(P\). Are there nontrivial bounds on the size of \(A\)?

Before stating Gallagher’s theorem, we require a few preliminaries. Recall that the von Mangoldt function \(\Lambda\) is defined by \(\Lambda(p^n) = \log p\) if \(p\) is a prime and \(n \geq 1\) and \(\Lambda(x) = 0\) for any other integer \(x\). The crucial property of \(\Lambda\) is that

\[
\sum_{d \mid n} \Lambda(d) = \log n
\]

for all \(n\), as it immediately follows from the definition and the unique factorization theorem.

**Theorem 5.** (Gallagher’s larger sieve) Let \(S\) be a finite nonempty set of integers and let \(P\) be a finite set of prime powers. Assume that for each \(q \in P\) we can find a real number \(u(q) \geq |\{s \pmod{q}|s \in S\}|\) such that

\[
\sum_{q \in P} \frac{\Lambda(q)}{u(q)} > \log 2X,
\]

where \(X \leq \max_{s \in S} |s|\). Then

\[
|S| \leq \frac{\sum_{q \in P} \Lambda(q) - \log 2X}{\sum_{q \in P} \frac{\Lambda(q)}{u(q)} - \log 2X}.
\]
Proof. Let \( q \in P \) and let \( s(r, q) \) be the number of elements of \( S \) that are congruent to \( r \) modulo \( q \). Then by Cauchy-Schwarz and the fact that \( u(q) \geq |\{ s \mod q | s \in S \}| \) we have

\[
|S|^2 = \left( \sum_{r=0}^{q-1} s(r, q) \right)^2 \leq u(q) \sum_{r=0}^{q-1} s(r, q)^2,
\]

thus

\[
\frac{|S|^2}{u(q)} \leq \sum_{r=0}^{q-1} \sum_{s_1, s_2 \in S, s_1 \equiv s_2 \equiv r \mod q} 1 \leq |S| + \sum_{s_1 \neq s_2 \in S} 1_{q|s_1 - s_2}.
\]

Multiplying this by \( \Lambda(q) \) and summing over all \( q \in P \) yields

\[
|S|^2 \sum_{q \in P} \frac{\Lambda(q)}{u(q)} \leq |S| \sum_{q \in P} \Lambda(q) + \sum_{s_1 \neq s_2} \sum_{q | s_1 - s_2} \Lambda(q).
\]

As

\[
\sum_{q | s_1 - s_2} \Lambda(q) \leq \log(|s_1 - s_2|) \leq \log 2X,
\]

we deduce that

\[
|S|^2 \sum_{q \in P} \frac{\Lambda(q)}{u(q)} \leq |S| \sum_{q \in P} \Lambda(q) + (|S|^2 - |S|) \log 2X,
\]

from which the result follows immediately. \( \square \)

We are now ready to prove that \( \log a \) and \( \log b \) are linearly dependent. Replacing \( a \) and \( b \) by \( a^2 \) and \( b^2 \), we may assume that \( a > 2 \). Suppose that \( a \) and \( b \) are \( \mathbb{Q} \)-linearly independent and consider a large integer \( x \). Let

\[
S_x = \{1, 2, \ldots, x\} \cap \{a^i b^j | i, j \in \mathbb{N}\}.
\]

On the one hand, by assumption \( a^i b^j = a^{i'} b^{j'} \) forces \( i = i' \) and \( j = j' \), so that \( |S_x| \) is the number of pairs \((i, j)\) for which \( i, j \geq 0 \) and \( i \log a + j \log b \leq \log x \). Hence there is an absolute constant \( c > 0 \) depending only on \( a \) and \( b \) such that \( |S_x| > c(\log x)^2 \) for all large enough \( x \).

On the other hand, we will use Gallagher’s sieve to prove that \( |S_x| \leq c' \log x \) for some absolute constant \( c' \) and all large enough \( x \). This will contradict the result established in the first paragraph and will end the proof of theorem 1.

For each positive integer \( y \) we let \( P_y \) be the set of prime powers \( q \) dividing at least one of the numbers \( a-1, a^2-1, \ldots, a^y-1 \). If \( q \in P_y \), let \( u(q) \) be the order of \( a \) mod \( q \). Since by assumption \( b \) is a power of \( a \) mod \( q \), it follows that \( \{ s \mod q | s \in S_x \} \) is bounded above by \( u(q) \) for all \( q \in P_y \). Also, note that \( u(q) \leq y \) for all \( q \in P_y \). The main technical point of the proof is the following:
Proposition 6. There are constants \(c_1, c_2 > 0\) depending only on \(a\) such that for all \(y \geq 1\) we have
\[
c_1 y^2 \leq \sum_{q \in P_y} \Lambda(q) \leq c_2 y^2.
\]

The proof of this proposition will occupy the rest of this note and is quite technical. Let us assume that the result holds and see how we can finish the proof of theorem 1. Choose \(y = [c \log x]\) for some suitable constant \(c\), such that
\[
\sum_{q \in P_y} \frac{\Lambda(q)}{u(q)} > 2 \log(2x).
\]
This is possible by the previous proposition and the fact that \(u(q) \leq y\) for all \(q \in P_y\). Using Gallagher’s sieve and the previous proposition again, we deduce the existence of an absolute constant \(c'\) for which
\[
|S_x| < c' \log x
\]
for all \(x\) large enough.

It remains to prove the proposition. The upper bound is quite easy, since
\[
\sum_{q \in P_y} \Lambda(q) \leq \sum_{j=1}^y \sum_{d|a_j-1} \Lambda(d) = \sum_{j=1}^y \log(a_j-1) < y^2 \log a.
\]
The lower bound is trickier and requires some standard results in the theory of cyclotomic polynomials. Let \(\Phi_n\) be the \(n\)th cyclotomic polynomial. Recall that \(\Phi_n(X)\) is the monic polynomial of degree \(\varphi(n)\) whose roots are the primitive roots of unity of order \(n\). In particular, we have \(\Phi_n(a) = |\Phi_n(a)| \geq (a-1)^{\varphi(d)}\) for all \(a \geq 2\). The fundamental property of \(\Phi_n\) is that for all primes \(p\) and all \(a > 1\), if \(p | \Phi_n(a)\) then \(p | n\) or the order of \(a \mod p\) is \(n\). It follows immediately from this that if \(q\) is a power of a prime and \(q | \phi_n(a)\), then \(\gcd(q,n) \neq 1\) or \(u(q) = n\). Hence
\[
\sum_{q \in P_y} \Lambda(q) \geq \sum_{d=1}^y \left( \sum_{q | \Phi_d(a)} \Lambda(q) - \sum_{q | d} \Lambda(q) \right) = \sum_{d=1}^y (\log \Phi_d(a) - \log d)
\]
and the previous discussion combined with \(\sum_{d=1}^y \log d < y \log y\) yields
\[
\sum_{q \in P_y} \Lambda(q) > \log(a-1) \cdot \sum_{d=1}^y \varphi(d) - y \log y.
\]
We conclude the proof of the proposition using the following classical result:

Proposition 7. For \(x \to \infty\) we have
\[
\sum_{k=1}^x \varphi(k) = \frac{3}{\pi^2} x^2 + O(x \log x).
\]
Proof. For the convenience of the reader, we give a sketch of the proof. Adding the relations \( \frac{\varphi(k)}{k} = \sum_{d|k} \frac{\mu(k)}{k} \) (where \( \mu \) is Mobius’ function) for \( k = 1, \ldots, x \) gives the relation

\[
\sum_{k=1}^{x} \frac{\varphi(k)}{k} = \sum_{d=1}^{x} \mu(d) \cdot \left( \frac{x}{d} \right) \left( 1 + \frac{x}{d} \right) = \frac{x^2}{2} \sum_{d=1}^{x} \frac{\mu(d)}{d^2} + O(x \log x).
\]

We conclude using the fact that \( \sum_{d \geq 1} \frac{\mu(d)}{d^2} = \frac{6}{\pi^2} \) and the estimate

\[
\left| \sum_{d=1}^{x} \frac{\mu(d)}{d^2} - \sum_{d \geq 1} \frac{\mu(d)}{d^2} \right| < \sum_{d>x} \frac{1}{d^2} = O\left( \frac{1}{x} \right).
\]

The relation \( \sum_{d \geq 1} \frac{\mu(d)}{d^2} = \frac{6}{\pi^2} \) follows by combining Euler’s classical formula \( \sum_{d \geq 1} \frac{1}{d^2} = \frac{\pi^2}{6} \) and the equality \( \sum_{d \mid n} \mu(d) = 0 \) for \( n > 1 \). These yield

\[
\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} \cdot \sum_{n \geq 1} \frac{1}{n^2} = \sum_{m \geq 1} \sum_{d \mid m} \frac{\mu(d)}{m^2} = 1.
\]

\[ \square \]

References


