

Heilbronn's Exponential Sum and Transcendence Theory

D.R. Heath-Brown
Mathematical Institute, Oxford

*To Professor Alan Baker F.R.S,
on the occasion of his sixtieth birthday*

Let p be a prime, and set $e(x) = \exp(2\pi ix)$. Heilbronn's exponential sum is defined to be

$$S(a, p) = \sum_{n=1}^p e\left(\frac{an^p}{p^2}\right),$$

for any integer a coprime to p . Although the sum appears to be defined modulo p^2 , one may observe that if $n \equiv n' \pmod{p}$, then $n^p \equiv n'^p \pmod{p^2}$. Thus the summand in $S(a, p)$ in fact has period p with respect to n . Heilbronn's sum is therefore a 'complete sum' to modulus p .

Heilbronn asked whether $S(a, p) = o(p)$ as $p \rightarrow \infty$. Methods based on algebraic geometry, in the spirit of Weil or Deligne, appear to be ineffectual for this problem, and elementary techniques have also failed to provide an answer. None the less we can now answer Heilbronn's question with the following theorem.

Theorem *If p is a prime and $p \nmid a$ then $S(a, p) \ll p^{7/8}$, uniformly in a .*

This result is due to Heath-Brown and Konyagin [2], there being an earlier estimate, due to Heath-Brown [1], with an exponent $11/12$.

To prove the theorem one begins with some elementary manipulations using the sum

$$S_0(a) = \sum_{n=1}^{p-1} e\left(\frac{an^p}{p^2}\right).$$

Since $S_0(a) = S_0(am^p)$ when $p \nmid m$ it follows that

$$(p-1) \sum_{r=1}^p |S_0(a+rp)|^4 = \sum_{r=1}^p \sum_{m=1}^{p-1} |S_0((a+rp)m^p)|^4 \leq \sum_{n=1}^{p^2} |S_0(n)|^4,$$

because each value of n arises at most once. We deduce that

$$(p-1) \sum_{r=1}^p |S_0(a+rp)|^4 \leq \sum_{m_1, \dots, m_4=1}^{p-1} \sum_{n=1}^{p^2} e_{p^2}((m_1^p + m_2^p - m_3^p - m_4^p)n) \\ = p^2 \# \{1 \leq m_1, \dots, m_4 \leq p-1 : m_1^p + m_2^p \equiv m_3^p + m_4^p \pmod{p^2}\}.$$

The final congruence implies that $m_1 + m_2 \equiv m_3 + m_4 \pmod{p}$, and hence $m_1 - m_3 \equiv m_4 - m_2 \equiv b \pmod{p}$, say. The case $p|b$ makes a negligible contribution. When $p \nmid b$ we write $m_1 \equiv v_1 b \pmod{p}$, whence $m_3 \equiv (v_1 - 1)b \pmod{p}$. Thus

$$m_1^p - m_3^p \equiv (v_1^p - (v_1 - 1)^p)b^p \pmod{p^2}.$$

Similarly we find that

$$m_4^p - m_2^p \equiv (v_2^p - (v_2 - 1)^p)b^p \pmod{p^2},$$

where $m_4 \equiv v_2 b \pmod{p}$.

The congruence $m_1^p + m_2^p \equiv m_3^p + m_4^p \pmod{p^2}$ now produces

$$(v_1^p - (v_1 - 1)^p)b^p \equiv (v_2^p - (v_2 - 1)^p)b^p \pmod{p^2}.$$

Since

$$v^p - (v-1)^p = \sum_{l=1}^p (-1)^{l-1} v^{p-l} \binom{p}{l} \equiv 1 - pf(v) \pmod{p^2},$$

it then follows, on allowing for the various possibilities for b , that

$$\begin{aligned} & \# \{1 \leq m_1, \dots, m_4 \leq p-1 : m_1^p + m_2^p \equiv m_3^p + m_4^p \pmod{p^2}\} \\ & \leq (p-1)^2 + (p-1) \# \{2 \leq v_1, v_2 \leq p-1 : f(v_1) \equiv f(v_2) \pmod{p}\}, \end{aligned}$$

where

$$f(X) = X + \frac{X^2}{2} + \frac{X^3}{3} + \dots + \frac{X^{p-1}}{p-1} \in \mathbb{Z}_p[X].$$

Thus

$$(p-1) \sum_{r=1}^p |S_0(a+rp)|^4 \leq p^2 \{(p-1)^2 + (p-1) \sum_{r=1}^p N_r^2\},$$

where N_r is the number of solutions $k \in \mathbb{Z}_p - \{0, 1\}$ of $f(k) = r$. This suffices for the following result.

Lemma 1 *We have*

$$S(a, p) \ll p^{1/2} \left\{ \sum_{r=1}^p N_r^2 \right\}^{1/4}.$$

Trivially one has $\sum_{r=1}^p N_r = p - 2$ and hence $\sum_{r=1}^p N_r^2 \ll p^2$. Since this leads to the estimate $S(a, p) \ll p$, we see that nothing has been lost up to this point. On the other hand, it is not so clear how any non-trivial estimate for N_r may be obtained.

It turns out that ideas from the work of Stepanov [4] are the key to handling N_r . Stepanov established Weil's theorem on the number of points on a curve over a finite field. However, his ideas can be applied to bound the number of zeros of a polynomial in one variable. A simple bound for N_r was obtained in this way by Mit'kin [3]. One begins by constructing an auxilliary polynomial

$$\Phi(X, Y, Z) \in \mathbb{Z}_p[X, Y, Z]$$

such that

$$\Psi(X) = \Phi(X, f(X), X^p)$$

vanishes to high order at roots of $f(X) = r$. This is a principle familiar from transcendence theory. Indeed the link goes much further, for the similarity between $f(X)$ and the function

$$-\log(1 - X) = X + \frac{X^2}{2} + \frac{X^3}{3} + \dots \in \mathbb{Q}[[X]]$$

is of crucial importance in the details of the argument. Thus the construction of the auxilliary polynomial $\Phi(X, Y, Z)$ depends on the fact that $f(X)$ satisfies some simple differential equations. These are given by the following result.

Lemma 2 *For any positive integer r there exist polynomials $q_r(X)$ and $h_r(X)$ in $\mathbb{Z}_p[X]$, of degrees at most r and $r - 1$ respectively, such that*

$$\{X(1 - X)\}^r \left(\frac{d}{dx} \right)^r f(X) = q_r(X) + (X^p - X)h_r(X).$$

Thus, although $f(X)$ has large degree, its derivatives may, in effect, be replaced by $q_r(X)$, which has small degree.

It is essential for the proof that $\Psi(X)$ should not vanish identically. Again this is a familiar aspect of transcendence arguments. For our situation we are motivated by the fact that $-\log(1 - X)$ is a transcendental function, and hence cannot satisfy a polynomial relation. Since $f(X)$ is almost equal to $-\log(1 - X)$ we expect that $f(X)$ similarly should not satisfy a polynomial relation of small degree. In fact one has the following result.

Lemma 3 *Let $F(X, Y) \in \mathbb{Z}_p[X, Y]$ have degree less than A with respect to X , and degree less than B with respect to Y . Then if F does not vanish identically we will have $X^p \nmid F(X, f(X))$, providing only that $AB < p$.*

As soon as $AB \geq p$, the polynomial F will have enough coefficients to ensure that $X^p \mid F(X, f(X))$ is possible. Thus the above result is surprisingly sharp.

In the author's work [1] Stepanov's method was applied in a simple-minded way, to show that $N_r = O(p^{2/3})$. This result had in fact been obtained earlier by Mit'kin [3]. Using Lemma 1, the above bound for N_r immediately produces the estimate $S(a, p) \ll p^{11/12}$. However in the later paper [2], the auxilliary polynomial was constructed so as to vanish for the roots of several different equations $f(X) = r$, thereby producing a bound for a sum

$$\sum_{r \in \mathcal{R}} N_r.$$

This leads to the superior exponent $7/8$ quoted in our theorem.

Two questions naturally arise. Firstly:- The sum $\sum_r N_r^2$ counts points on the curve $f(X) = f(Y)$. Is there a way of attacking this directly, rather than handling individual values of N_r ? Secondly:- The function $-\log(1 - X)$ satisfies a first order differential equation. Can one handle problems in which the function corresponding to $f(X)$ is related to a solution of a second (or higher) order equation?

References

- [1] D.R. Heath-Brown, An estimate for Heilbronn's exponential sum, *Analytic number theory: Proceedings of a conference in honor of Heini Halberstam*, (Birkhäuser, Boston, 1996), 451-463.
- [2] D.R. Heath-Brown, and S Konyagin, New bounds for Gauss sums derived from k -th powers, and for Heilbronn's exponential sum, *Quart. J. Math. Oxford Ser.*, (to appear).
- [3] D.A. Mit'kin, An estimate for the number of roots of some comparisons by the Stepanov method, *Mat. Zametki*, 51 (1992), 52-58, 157. (Translated as *Math. Notes*, 51 (1992), 565-570.)
- [4] S.A. Stepanov, The number of points of a hyperelliptic curve over a prime field, *Izv. Akad. Nauk SSSR Ser. Mat.*, 33 (1969), 1171-1181.