On the Distribution of Pseudopowers

Sergei V. Konyagin
Department of Mechanics and Mathematics
Moscow State University
Moscow, 119992, Russia
konyagin@ok.ru

Carl Pomerance
Department of Mathematics
Dartmouth College
Hanover, NH 03755-3551, USA
carlp@gauss.dartmouth.edu

Igor E. Shparlinski
Department of Computing
Macquarie University
Sydney, NSW 2109, Australia
igor@ics.mq.edu.au

Abstract

An x-pseudopower to base g is a positive integer which is not a power of g yet is so modulo p for all primes $p \le x$. We improve an upper bound for the least such number due to E. Bach, R. Lukes, J. Shallit, and H. C. Williams. The method is based on a combination of some bounds of exponential sums with new results about the average behaviour of the multiplicative order of g modulo prime numbers.

1 Introduction

Let g be a fixed integer with $|g| \geq 2$. Following E. Bach, R. Lukes, J. Shallit, and H. C. Williams [1], we say that an integer $n > 0$ is an *x-pseudopower to base g* if n is not a power of g over the integers but is a power of g modulo all primes $p \leq x$, that is, if for all primes $p \leq x$ there exists an integer $e_p \geq 0$ such that $n \equiv g^{e_p} \pmod{p}$.

Denote by $q_g(x)$ the least x-pseudopower to base g.

A well-known result of A. Schinzel [20] asserts that if f and $g > 0$ are integers, such that $f \neq g^k$ for all integers $k \geq 0$, then for infinitely many primes p the congruence $g^x \equiv f \pmod{p}$ does not have solutions in nonnegative integers x. Therefore,

$$q_g(x) \to \infty, \qquad x \to \infty.$$

E. Bach, R. Lukes, J. Shallit and H. C. Williams [1] have shown that if the Riemann hypothesis holds for Dedekind zeta functions, then there is a constant $A > 0$, depending only on g, such that

$$q_g(x) \geq \exp(A\sqrt{x}/(\log x)^2).$$

On the other hand, if

$$M_x = \prod_{p \leq x} p$$

is the product of all primes $p \leq x$, then $q_g(x) \leq 2M_x + 1$ when $x \geq 2$. Indeed, both $M_x + 1$ and $2M_x + 1$ are $\equiv g^0 \pmod{p}$ for all primes $p \leq x$ and evidently not both can be powers of g. The prime number theorem implies that $M_x = e^{(1+o(1))x}$, so we have

$$q_g(x) \leq e^{(1+o(1))x}, \qquad x \to \infty. \tag{1}$$

Though the inequality $q_g(x) \leq 2M_x + 1$ cannot be improved in general (consider the case $g = M_x + 1$), if g is fixed or $|g|$ is not too large compared with x, there is a chance to improve the bound (1). Supported by numerical data, a heuristic argument is given in [1] suggesting that $q_g(x)$ for fixed g is about $\exp(c_g x/\log x)$, where $c_g > 0$. We obtain a more modest upper bound valid for $|g| \leq x$ as well as several more results about the distribution of x-pseudopowers to base g.

For an integer m we use $\mathbb{Z}_m$ to denote the residue ring modulo m. Now, for a prime p, we denote by $\mathcal{U}_{g,p}$ the subset of $\mathbb{Z}_p$ generated by powers of g modulo p, that is

$$\mathcal{U}_{g,p} = \{n \in \mathbb{Z}_p \ : \ n \equiv g^k \pmod p \ \text{ for some nonnegative } k \in \mathbb{Z}\}.$$

Clearly, if $\gcd(g,p) = 1$ then $\mathcal{U}_{g,p}$ is a subgroup of $\mathbb{Z}_p^*$, while if $p \mid g$, then $\mathcal{U}_{g,p} = \{0,1\}$.

We consider the set

$$\mathcal{W}_g(x) = \{n \in [0, M_x) \ : \ n \in \mathcal{U}_{g,p} \text{ for all primes } p \le x\}.$$

The set $\mathcal{W}_g(x)$ consists of both the x-pseudopowers to base g that lie below M_x and the true powers of g in this range. (In the case that $M_x \mid g$, the set $\mathcal{W}_g(x)$ also contains 0, but we assume that $|g| \le x$ and x is large, so that this case does not occur.) The number of true powers of g below M_x is $O(x)$, which turns out to be minuscule in comparison to $\#\mathcal{W}_g(x)$.

We first get a good lower bound for $\#\mathcal{W}_g(x)$. Then we estimate exponential sums with elements of $\mathcal{W}_g(x)$ and use these bounds to derive some uniformity-of-distribution results for elements of $\mathcal{W}_g(x)$. Our estimate for $q_g(x)$ follows from these results.

2 Our approach and results

Our approach is based on a combination of two techniques:

- recent bounds of exponential sums over reasonably small subgroups of the multiplicative group $\mathbb{Z}_p^*$ due to Heath-Brown and Konyagin [10];
- Lower bounds on multiplicative orders on average which we derive from upper bounds of R. C. Baker and G. Harman [2, 3] (which are summarised in [9]) for the Brun–Titchmarsh inequality on average.

We do not try to obtain numerically the best results, rather we concentrate on the exposition of our main ideas. Certainly with more work and numerical calculations one can get more precise results. Furthermore, any further advance in our knowledge on the above two topics would immediately lead to further progress on this problem as well.

For prime $p \nmid g$, let $l_g(p) = \#\mathcal{U}_{g,p}$, the multiplicative order of g modulo p. We also put $l_g(p) = 1$ for $g \equiv 0 \pmod p$. We now define the product

$$R_g(x) = \prod_{p \le x} l_g(p). \tag{2}$$

The Chinese remainder theorem implies that

$$\#\mathcal{W}_g(x) = \prod_{p \le x} \#\mathcal{U}_{g,p}.$$

Further, for $p \mid g$, we have $\#\mathcal{U}_{g,p} = 2 = 2l_g(p)$. Thus, if $\gcd(g, M_x)$ has exactly k prime factors,

$$\#\mathcal{W}_g(x) = 2^k R_g(x) \ge R_g(x). \tag{3}$$

Note that $R_g(x)^{1/\pi(x)}$ is the geometric mean of $l_g(p)$ for $p \le x$ and so has some independent interest. Our first result gives a lower bound for $R_g(x)$ and so, via (3), gives a lower bound for $\#\mathcal{W}_g(x)$.

Theorem 1. *For x sufficiently large and for g an integer with $2 \le |g| \le x$, we have*

$$\#\mathcal{W}_g(x) \ge R_g(x) \ge \exp(\eta x)$$

where $\eta = 0.58045$.

We put $\mathbf{e}(u) = \exp(2\pi i u)$ and define exponential sums

$$S_{a,g}(x) = \sum_{n \in \mathcal{W}_g(x)} \mathbf{e}(an/M_x).$$

Theorem 2. *For x sufficiently large and for any integers a, g with $2 \le |g| \le x$, we have*

$$|S_{a,g}(x)| \le \#\mathcal{W}_g(x) \gcd(a, M_x) \exp(-\gamma x)$$

where

$$\gamma = 0.11286.$$

For a positive integer $h \le M_x$, let $N_g(x,h)$ denote the number of members of $\mathcal{W}_g(x)$ below h. Using some standard arguments, we derive from our estimates of the sums $S_{a,g}(x)$:

Theorem 3. *For x sufficiently large, we have for any integers g and h with $2 \le |g| \le x$ and $1 \le h \le M_x$,*

$$N_g(x,h) = \#\mathcal{W}_g(x)\frac{h}{M_x} + E_g(x,h)$$

where

$$|E_g(x,h)| \le \#\mathcal{W}_g(x)\exp(-\gamma x)$$

and where γ is as in Theorem 2.

In particular, we improve (1) to

$$q_g(x) \le e^{0.88715x}$$

for x sufficiently large and $|g| \le x$. Indeed, if we take $h = e^{0.88715x}$ in Theorem 3, then that result implies that there are at least $\frac{1}{2}\#\mathcal{W}_g(x)h/M_x$ numbers in $\mathcal{W}_g(x)$ below h. Together with Theorem 1 this implies that there are more than $e^{.4675x}$ members of $\mathcal{W}_g(x)$ below h. But there are only $O(x)$ numbers below h that are true powers of g, so there are many members of $\mathcal{W}_g(x)$ below h that are x-pseudopowers to base g.

3 Proof of Theorem 1

It is well known, see [5, 6, 12, 17], that $l_g(p) \ge x^{1/2}$ for all but $o(x/\log x)$ primes $p \le x$. Thus for $R_g(x)$, given by (2), we immediately obtain

$$R_g(x) \ge \exp(x/2 + o(x)). \qquad (4)$$

We now obtain a more accurate estimate for $R_g(x)$.

Let $P(m)$ denote the largest prime divisor of $m \ge 2$ (with the convention $P(1) = 0$). We use $\pi(x,y)$ to denote the number of primes $p \le x$ with $P(p-1) \le y$ and define the constant

$$c = \liminf \pi(x, x^{1/2})/\pi(x). \qquad (5)$$

Lemma 4. *For the product $R_g(x)$, given by (2), we have*

$$R_g(x) \ge \exp\left(\frac{1+c}{2}x + o(x)\right),$$

where c is given by (5).

Proof. Let $\mathcal{P}_0$ be the set of primes $p \le x$ with $l_g(p) \le x^{1/2}$, let $\mathcal{P}_1$ be the set of primes $p \le x$ with $l_g(p) > x^{1/2}$ and $P(p-1) > x^{1/2}$, and let $\mathcal{P}_2$ be the set of all other primes $p \le x$.

We simply ignore the contribution from primes in $\mathcal{P}_0$ (which, as we have mentioned, is $\exp(o(x))$ anyway).

For each $p \in \mathcal{P}_1$, since $l_g(p) \mid p-1$ and $(p-1)/P(p-1) < x^{1/2}$, we have $P(p-1) \mid l_g(p)$. Thus,

$$\sum_{p\in\mathcal{P}_1} \log l_g(p) \ge \sum_{p\in\mathcal{P}_1} \log P(p-1) = \sum_{x^{1/2}<q\le x} \pi(x;q,1)\log q \ + o(x), \tag{6}$$

where q runs over primes and $\pi(x;k,b)$ denotes the number of primes $p \le x$ with $p \equiv b \pmod{k}$. Indeed, each q in the indicated range corresponds to $\pi(x;q,1)$ primes $p \le x$ with $P(p-1) = q$, and almost all primes p so counted in the sum are in $\mathcal{P}_1$. It follows from the Bombieri–Vinogradov theorem and the Brun–Titchmarsh inequality (see [13, Theorems 6.6 and 17.1]) that

$$\sum_{q\le x^{1/2}} \pi(x;q,1)\log q = (1/2+o(1))x,$$

and since $\sum_{q\le x} \pi(x;q,1)\log q = (1+o(1))x$, we have

$$\sum_{x^{1/2}<q\le x} \pi(x;q,1)\log q = (1/2+o(1))x, \tag{7}$$

as noted by M. Goldfeld [8]. We thus have from (6) that

$$\sum_{p\in\mathcal{P}_1} \log l_g(p) \ge (1/2+o(1))x. \tag{8}$$

We now consider the contribution from primes in $\mathcal{P}_2$. For each such prime p we have $l_g(p) \ge x^{1/2}$, so that

$$\sum_{p\in\mathcal{P}_2} \log l_g(p) \ge \frac{1}{2}\log x \sum_{p\in\mathcal{P}_2} 1 = \frac{1}{2}\pi(x, x^{1/2})\log x + o(x). \tag{9}$$

The bounds (8) and (9), together with (5), imply that

$$\sum_{p\le x} \log l_g(p) \ge (1/2 + c/2 + o(1))x,$$

which concludes the proof. □

There is probably little doubt that

$$c = \rho(2) = 1 - \log 2 = 0.3068\ldots,$$

where $\rho(u)$ is the Dickman–de Bruijn function (see [21]), however proving this seems to be inaccessible by present methods; see [4, 18, 19] where more general conjectures about $\pi(x, y)$ are discussed. Note that in [18] we have the inequality

$$\pi(x, x^{1/2}) \geq (1 - 4\log(5/4) + o(1))x/\log x,$$

so that $c \geq 0.107425\ldots$. The key tool in [18] is a result of C. Hooley [11] from 1973. Using more modern tools we now obtain a larger value of c.

For $1/2 \leq u < 1$ let $C(u)$ denote a monotone nondecreasing function such that for any $\varepsilon > 0$ and $A > 0$, we have

$$\pi(x; k, b) \leq (C(u) + \varepsilon)\frac{x}{\varphi(k)\log x} \tag{10}$$

for all integers $k \leq x^u$ but for at most $x^u/\log^A x$ exceptions, for all b coprime to k for allowable values of k, and for all $x \geq x_0(A, \varepsilon)$. H. L. Montgomery and R. C. Vaughan [16] have a version of the Brun–Titchmarsh theorem which allows one to take $C(u) = 2/(1-u)$ with no exceptional values of k and with $\varepsilon = 0$, see also [9, Theorem 8.1] or [13, Section 6.8]). But allowing a small exceptional set as indicated here then permits one to get smaller values of $C(u)$. This is the arena of "the Brun–Titchmarsh theorem on average." The key results we use are due to É. Fouvry [7] and R. C. Baker and G. Harman [2, 3]. (There are many other contributors to this subject, we refer to [9] for more details and further references).

For a monotone nondecreasing function $C(u)$ satisfying (10), let us define ϑ_C by the equation

$$\int_{1/2}^{\vartheta_C} C(u)\,du = 1/2. \tag{11}$$

(Note that for any monotone nondecreasing function $C(u)$ the integral is well defined.)

We now use the approach of [18] to show the following lower bound on c.

Lemma 5. *For the constant c given by* (5), *we have*

$$c \geq 1 - \int_{1/2}^{\vartheta_C} \frac{C(u)}{u}\,du$$

where $C(u)$ is an arbitrary monotone nondecreasing function satisfying (10) *and ϑ_C is defined by* (11).

Proof. Let

$$H(x,t) = \sum_{x^{1/2}<q\le t} \pi(x;q,1)\log q$$

where q runs over primes. Thus, by partial summation, we have

$$\pi(x) - \pi(x,x^{1/2}) = \sum_{x^{1/2}<q\le x} \pi(x;q,1) = \frac{H(x,x)}{\log x} + \int_{x^{1/2}}^{x} \frac{H(x,t)}{t\log^2 t}\,dt. \qquad (12)$$

Using (7), the first term on the right in (12) is $(1/2+o(1))x/\log x$, so it remains to get a good upper bound for the integral.

Using the inequality (10), partial summation, and the prime number theorem, we have

$$H(x,t) \le x\int_{1/2}^{\log t/\log x} C(u)\,du + o(x). \qquad (13)$$

Thus, for any value of $\vartheta \in (1/2,1)$ we have

$$\int_{x^{1/2}}^{x^{\vartheta}} \frac{H(x,t)}{t\log^2 t}\,dt \le x\int_{x^{1/2}}^{x^{\vartheta}} \frac{1}{t\log^2 t}\int_{1/2}^{\log t/\log x} C(u)\,du\,dt + o(x/\log x). \qquad (14)$$

By a change of variables and an interchange of the order of integration, the double integral is equal to

$$\begin{aligned}
\int_{x^{1/2}}^{x^{\vartheta}} \frac{1}{t\log^2 t}\int_{1/2}^{\log t/\log x} C(u)\,du\,dt &= \int_{1/2}^{\vartheta} C(u)\int_{x^u}^{x^{\vartheta}} \frac{1}{t\log^2 t}dt\,du \\
&= \int_{1/2}^{\vartheta} C(u)\int_{u\log x}^{\vartheta\log x} \frac{1}{v^2}dv\,du \\
&= \frac{1}{\log x}\int_{1/2}^{\vartheta} C(u)\left(\frac{1}{u}-\frac{1}{\vartheta}\right)\,du.
\end{aligned}$$

Thus, from (14) we have

$$\int_{x^{1/2}}^{x^{\vartheta}} \frac{H(x,t)}{t\log^2 t}\,dt \le \frac{x}{\log x}\int_{1/2}^{\vartheta} C(u)\left(\frac{1}{u}-\frac{1}{\vartheta}\right)\,du + o(x/\log x).$$

Using $H(x,t) \le H(x,x) = (1/2 + o(1))x$ (see (7)), we then have for any $\vartheta \in (1/2, 1)$ that

$$\begin{aligned}\int_{x^{1/2}}^{x} \frac{H(x,t)}{t\log^2 t}\,dt &= \int_{x^{1/2}}^{x^{\vartheta}} \frac{H(x,t)}{t\log^2 t}\,dt + \int_{x^{\vartheta}}^{x} \frac{H(x,t)}{t\log^2 t}\,dt \\ &\le \frac{x}{\log x}\int_{1/2}^{\vartheta} C(u)\left(\frac{1}{u} - \frac{1}{\vartheta}\right)du + \frac{x}{2\log x}\left(\frac{1}{\vartheta} - 1 + o(1)\right)\end{aligned}$$

which we rewrite as

$$\begin{aligned}&\int_{x^{1/2}}^{x} \frac{H(x,t)}{t\log^2 t}\,dt \\ &\quad\le \frac{x}{\log x}\left(\int_{1/2}^{\vartheta} \frac{C(u)}{u}\,du - \frac{1}{\vartheta}\int_{1/2}^{\vartheta} C(u)du + \frac{1}{2\vartheta} - \frac{1}{2} + o(1)\right).\end{aligned} \tag{15}$$

If we choose $\vartheta = \vartheta_C$ defined by (11), then using (12) and (15), we obtain

$$\pi(x) - \pi(x, x^{1/2}) \le \frac{x}{\log x}\int_{1/2}^{\vartheta_C} \frac{C(u)}{u}\,du + o(x/\log x)$$

which concludes the proof. □

We now use known results on the possible choices of the function $C(u)$ in (10), as summarised in [9], to obtain a lower bound for c.

Lemma 6. *For the constant c given by* (5), *we have*

$$c > 0.160901.$$

Proof. For $u \in [0.51, 0.56]$, we define $C(u)$ as a step-wise monotonically nondecreasing function whose values at $u = 0.533$ and $u = 0.5 + 0.005j$, $j = 1, \ldots, 12$ are given by G. Harman in [9, Theorem 8.2] as $C(0.533) = 2$ and in [9, Table 8.1] as:

u	$C(u)$	u	$C(u)$	u	$C(u)$	u	$C(u)$	u	$C(u)$
0.515	1.223	0.525	1.75	0.535	2.09	0.545	2.47	0.555	2.76
0.52	1.632	0.53	1.82	0.540	2.25	0.55	2.66	0.56	2.88

For other values of u, we also use analytic expressions which are due to R. C. Baker and G. Harman [2, 3] and É. Fouvry [7]. These results are also presented in [9, page 184] (for $u \in [0.5, 0.51)$) and in [9, Theorem 8.4] ($u \in [17/32, 5/7]$):

- for $0.5 \leq u < 0.51$, we have $C(u) = 1 + 150(u - 1/2)^2$;
- for $17/32 < u \leq 4/7$, we have $C(u) = 14/(12 - 13u) - \log(4(1-u)/3u)$ (in fact we use it only for $0.56 < u \leq 4/7$);
- for $4/7 < u \leq 3/5$ we have $C(u) = 14/(12 - 13u)$;
- for $3/5 < u \leq 5/7$ we have $C(u) = 8/(3 - u)$.

With this we compute (using *Mathematica*)

$$\int_{0.5}^{0.51} C(u)\,du = 0.01005\,, \qquad \int_{0.51}^{0.56} C(u)\,du = 0.107405\,,$$
$$\int_{0.56}^{4/7} C(u)\,du \approx 0.034177\,, \qquad \int_{4/7}^{3/5} C(u)\,du \approx 0.091260\,,$$
$$\int_{3/5}^{.6759} C(u)\,du \approx 0.257087\,, \qquad \int_{3/5}^{.67591} C(u)\,du \approx 0.257121\,,$$

where the approximations are rounded to 6 decimal places.

Therefore

$$\int_{1/2}^{0.6759} C(u)\,du < 0.49999, \qquad \int_{1/2}^{0.67591} C(u)\,du > 0.50001,$$

and we see that for our choice of $C(u)$,

$$0.6759 < \vartheta_C < 0.67591.$$

We also compute

$$\int_{0.5}^{0.51} \frac{C(u)}{u}\,du < 0.019902\,, \qquad \int_{0.51}^{0.56} \frac{C(u)}{u}\,du < 0.199610\,,$$
$$\int_{0.56}^{4/7} \frac{C(u)}{u}\,du < 0.060412\,, \qquad \int_{4/7}^{3/5} \frac{C(u)}{u}\,du < 0.155787\,,$$
$$\int_{3/5}^{0.67591} \frac{C(u)}{u}\,du < 0.403388\,.$$

Therefore,

$$\int_{1/2}^{\vartheta_C} \frac{C(u)}{u}\,du < \int_{1/2}^{0.67591} \frac{C(u)}{u}\,du < 0.839099\,,$$

so that with Lemma 5 the result follows. □

Theorem 1 now follows from Lemmas 4 and 6.

4 Proof of Theorem 2

Let

$$T_g(x) = \prod_{\substack{p\le x\\ l_g(p)>p^{2/3}}} (l_g(p)p^{-2/3}). \tag{16}$$

We now show that the exponential sum $S_{a,g}(x)$ is related to $T_g(x)$ and the product $R_g(x)$ defined in (2).

Lemma 7. *For any integer a, we have*

$$|S_{a,g}(x)| \le \#\mathcal{W}_g(x) d \exp\left(x/4 + o(x)\right) R_g(x)^{-5/8} T_g(x)^{-3/8}$$

as $x \to \infty$, where $d = \gcd(a, M_x)$.

Proof. By the Chinese remainder theorem we see that

$$S_{a,g}(x) = \sum_{n\in\mathcal{W}_g(x)} \mathbf{e}(an/M_x) = \prod_{p\le x} \sum_{n\in\mathcal{U}_{g,p}} \mathbf{e}(a_p n/p) \tag{17}$$

where $a_p \in \mathbb{Z}_p$ is determined by the condition

$$a_p(M_x/p) \equiv a \pmod p.$$

If $p \nmid ag$, the bound of D. R. Heath-Brown and S. V. Konyagin [10] applies which gives the estimate

$$\left|\sum_{n\in\mathcal{U}_{g,p}} \mathbf{e}(a_p n/p)\right| \le C l_g(p)^{3/8} p^{1/4}. \tag{18}$$

for some absolute constant $C > 1$. We also recall the well-known bound

$$\left|\sum_{n\in\mathcal{U}_{g,p}} \mathbf{e}(a_p n/p)\right| \le p^{1/2} \tag{19}$$

(provided $p \nmid ag$), which is better than (18) for $l_g(p) > p^{2/3}$, see [14, Theorem 3.4].

For the set $\mathcal{P}_0$ of primes $p \le x$ with $p \mid ag$ we estimate the exponential sums over $\mathcal{U}_{g,p}$ trivially as $2l_g(p)$.

For the set $\mathcal{P}_1$ of primes with $p \nmid ag$ and $l_g(p) \le p^{2/3}$ we use the bound (18).

Finally, for the set $\mathcal{P}_2$ of primes with $p \nmid ag$ and $l_g(p) > p^{2/3}$ we use the bound (19).

Thus, substituting these bounds in (17), we obtain

$$|S_{a,g}(x)| \le 2^{\#\mathcal{P}_0} C^{\#\mathcal{P}_1} \prod_{p \in \mathcal{P}_0} l_g(p) \prod_{p \in \mathcal{P}_1} l_g(p)^{3/8} p^{1/4} \prod_{p \in \mathcal{P}_2} p^{1/2}. \tag{20}$$

We majorize the first two factors in (20) as $e^{O(\pi(x))} = e^{o(x)}$. The first product in (20) may be restricted to the primes $p \le x$ which divide a, and since $l_g(p) < p$, this product is bounded by $\gcd(a, M_x) = d$. Let $\mathcal{Q}_1, \mathcal{Q}_2$ be the same as $\mathcal{P}_1, \mathcal{P}_2$ but without the restriction that $p \nmid ag$. Thus, the three products in (20) are at most

$$\begin{aligned} d \prod_{p \in \mathcal{Q}_1} l_g(p)^{3/8} p^{1/4} \prod_{p \in \mathcal{Q}_2} p^{1/2} &= d \prod_{p \le x} l_g(p)^{3/8} p^{1/4} \prod_{p \in \mathcal{Q}_2} l_g(p)^{-3/8} p^{1/4} \\ &= d R_g(x)^{3/8} M_x^{1/4} T_g(x)^{-3/8}. \end{aligned}$$

Thus, the result follows from (20), the prime number theorem in the form $M_x = e^{(1+o(1))x}$, and the inequality (3). □

Using the elementary bound (4) together with Lemma 7 and the trivial bound $T_g(x) \ge 1$ already gives a nontrivial estimate on the sums $S_{a,g}(x)$, namely

$$|S_{a,g}(x)| \le \#\mathcal{W}_g(x) \gcd(a, M_x) \exp\left(-x/16 + o(x)\right).$$

Using Theorem 1 in place of (4) and still using only $T_g(x) \ge 1$ we get

$$|S_{a,g}(x)| \le \#\mathcal{W}_g(x) \gcd(a, M_x) \exp\left(-0.11278x\right)$$

for all large x. We now obtain a nontrivial estimate for $T_g(x)$, which in turn implies a slightly better estimate for $S_{a,g}(x)$.

Lemma 8. *For the product $T_g(x)$ given by* (16)*, a function $C(u)$ satisfying* (10)*, and $\vartheta_C > \frac{2}{3}$ defined by* (11) *we have*

$$T_g(x) \ge \exp\left(x \int_{2/3}^{\vartheta_C} \left(1 - \frac{2}{3u}\right) C(u)\, du + o(x)\right).$$

Proof. Let $\mathcal{P}$ be the set of primes $p \le x$ with $l_g(p) > x^{1/2}$ and $P(p-1) > x^{2/3}$. Similarly to the proof of Lemma 4 we have $l_g(p) > p^{2/3}$ for all $p \in \mathcal{P}$ and so

$$\begin{aligned}\log T_g(x) &\ge \sum_{p\in\mathcal{P}}(\log l_g(p) - \frac{2}{3}\log p)\\ &\ge \sum_{x^{2/3}<q\le x}\pi(x;q,1)(\log q - \frac{2}{3}\log x) + o(x),\end{aligned} \tag{21}$$

where q runs over primes.

Next, we follow the proof of Lemma 5. By partial summation, we have

$$\begin{aligned}&\sum_{x^{2/3}<q\le x}\pi(x;q,1)(\log q - \frac{2}{3}\log x)\\ &\qquad = \frac{1}{3}H(x,x) - \frac{2\log x}{3}\int_{x^{2/3}}^{x}\frac{H(x,t)}{t\log^2 t}\,dt.\end{aligned} \tag{22}$$

Using (13) as in the argument for (15), and recalling (11), we get

$$\begin{aligned}\int_{x^{2/3}}^{x}\frac{H(x,t)}{t\log^2 t}\,dt \;&\le\; \frac{x}{\log x}\Bigg(\int_{2/3}^{\vartheta_C}\frac{C(u)}{u}\,du + \frac{3}{2}\int_{1/2}^{2/3}C(u)du\\ &\qquad - \frac{1}{\vartheta_C}\int_{1/2}^{\vartheta_C}C(u)du + \frac{1}{2\vartheta_C} - \frac{1}{2} + o(1)\Bigg)\\ &= \frac{x}{\log x}\left(\int_{2/3}^{\vartheta_C}\frac{C(u)}{u}\,du + \frac{3}{2}\int_{1/2}^{2/3}C(u)du - \frac{1}{2} + o(1)\right).\end{aligned}$$

Combining this with (21) and (22), and then using (7), we complete the proof. □

Using the estimates for the function $C(u)$ as discussed in the proof of Lemma 6 we can now get an explicit estimate for $T_g(x)$.

Lemma 9. *For the product $T_g(x)$, given by* (16)*, and x sufficiently large, we have*

$$T_g(x) \ge \exp(0.000217x).$$

Proof. This follows immediately from Lemma 8, the estimate $\vartheta_C > 0.6759$ seen in the proof of Lemma 6, and the formula $C(u) = 8/(3-u)$ for the range $[3/5, 5/7]$ also seen in the proof of Lemma 6. □

We now have Theorem 2 by using, in the inequality of Lemma 7, our estimate for $R_g(x)$ in Theorem 1 and our estimate for $T_g(x)$ in Lemma 9.

5 Proof of Theorem 3

Using that for any integer $m \ge 1$ we have

$$\sum_{a=0}^{m-1} \mathbf{e}(au/m) = \begin{cases} 0, & \text{if } u \not\equiv 0 \pmod m, \\ m, & \text{if } u \equiv 0 \pmod m, \end{cases}$$

(which follows from the formula for the sum of a geometric progression) we write

$$N_g(x,h) = \sum_{n \in \mathcal{W}_g(x)} \sum_{k=0}^{h-1} \frac{1}{M_x} \sum_{a=0}^{M_x-1} \mathbf{e}\left(a(n-k)/M_x\right).$$

Changing the order of summation and separating the term $\#\mathcal{W}_g(x)h/M_x$ corresponding to $a=0$ we derive

$$\left| N_g(x,h) - \#\mathcal{W}_g(x)\frac{h}{M_x} \right| \le \frac{1}{M_x}\Delta \tag{23}$$

where

$$\Delta = \sum_{a=1}^{M_x-1} |S_{a,g}(x)| \left| \sum_{k=0}^{h-1} \mathbf{e}\left(ak/M_x\right) \right|.$$

For each $d \mid M_x$ with $d < M_x$ we now collect together the terms with $\gcd(a, M_x) = d$ and also apply Lemma 7, getting the estimate

$$\begin{aligned} \Delta &\le \#\mathcal{W}_g(x) \exp\left(x/4 + o(x)\right) R_g(x)^{-5/8} T_g(x)^{-3/8} \\ &\qquad\qquad \sum_{\substack{d < M_x \\ d \mid M_x}} d \sum_{\substack{a=1 \\ \gcd(a,M_x)=d}}^{M_x-1} \left| \sum_{k=0}^{h-1} \mathbf{e}\left(ak/M_x\right) \right| \\ &\le \#\mathcal{W}_g(x) \exp\left(x/4 + o(x)\right) R_g(x)^{-5/8} T_g(x)^{-3/8} \\ &\qquad\qquad \sum_{\substack{d < M_x \\ d \mid M_x}} d \sum_{b=1}^{M_x/d-1} \left| \sum_{k=0}^{h-1} \mathbf{e}\left(bk/(M_x/d)\right) \right|. \end{aligned}$$

We now recall that for any integers $m \ge 2$ and $1 \le b < m$, we have the bound

$$\left| \sum_{k=0}^{h-1} \mathbf{e}\left(bk/m\right) \right| \ll \frac{m}{\min\{b, m-b\}}$$

which again follows from the formula for the sum of a geometric progression, see [13, Bound (8.6)]. This implies that

$$\sum_{b=1}^{m-1}\left|\sum_{k=0}^{h-1}\mathbf{e}\left(bk/m\right)\right| \ll m\log m.$$

Thus

$$\Delta \le \#\mathcal{W}_g(x)M_x \exp\left(x/4+o(x)\right)R_g(x)^{-5/8}T_g(x)^{-3/8}$$

where we used that

$$\sum_{\substack{d<M_x\\ d\mid M_x}} 1 \le 2^{\pi(x)} = \exp\left(o(x)\right).$$

Substituting this bound in (23), we obtain

$$\left|N_g(x,h)-\#\mathcal{W}_g(x)\frac{h}{M_x}\right| \le \#\mathcal{W}_g(x)\exp\left(x/4+o(x)\right)R_g(x)^{-5/8}T_g(x)^{-3/8}.$$

Theorem 3 now follows from our estimates for $R_g(x)$ and $T_g(x)$ in Sections 3 and 4, respectively.

6 Remarks

Using better estimates for $C(u)$ that already exist, it is possible to get a larger value of ϑ_C and consequently better numbers in Lemmas 6 and 9. In particular in [3] and [9] a method of computing a somewhat smaller function C satisfying (10) is described leading to $\vartheta_C > 0.677$. Using this value of ϑ_C in our estimate for $T_g(x)$ allows us to replace 0.000217 with 0.000272. The changes in the estimate for c in Lemma 6 depend much more intrinsically on the better estimates for $C(u)$ that support a value of ϑ_C that is greater than 0.677; we have not worked this out.

Certainly if more information about the possible choice of the function $C(u)$ becomes available, one can immediately obtain even better numerical estimates for the constant c and thus improve the results of Theorems 1, 2, and 3.

Another avenue for improvement could come with our estimate for $l_g(p)$ when $P(p-1)\le\sqrt{x}$. We used the estimate $l_g(p)\ge\sqrt{x}$ for almost all such primes $p\le x$. It follows from [6, Theorem 6] of K. Ford that there is some $\varepsilon > 0$ such that for a positive proportion of these primes we have $l_g(p) \geq x^{1/2+\varepsilon}$. Having a version of this theorem with explicit constants would allow a numerical improvement in our Lemma 4 and thus an improvement in our principal results.

It is very plausible that the technique of [14, Chapter 7] can be used to improve our bound on $q_g(x)$ (but not the bounds of Theorems 2 and 3). However adjusting this technique to the case of composite moduli and then tuning it to accomodate in an optimal way our current knowledge of the behaviour of $l_g(p)$ may take significant efforts.

Finally, we recall that under the Generalised Riemann Hypothesis we have $l_g(p) = p^{1+o(1)}$ for almost all primes p, see [5, 15, 17], which immediately gives

$$R_g(x) = \exp(x + o(x)) \qquad \text{and} \qquad T_g(x) = \exp(x/3 + o(x)).$$

In turn, this means that one can take any $\gamma < 1/2$ in Theorems 2 and 3 and one has $q_g(x) \leq e^{x/2+o(x)}$.

Acknowledgments

We thank G. Harman for sending us a pre-publication file for Chapter 8 of his book [9]. S.V.K. gratefully acknowledges support from Grants 05-01-00066 from the Russian Foundation for Basic Research and NSh-5813.2006.1. C.P. gratefully acknowledges support from NSF grants DMS-0401422 and DMS-0703850. I.E.S. gratefully acknowledges support from ARC grant DP0556431.

References

[1] E. Bach, R. Lukes, J. Shallit and H. C. Williams, 'Results and estimates on pseudopowers', *Math. Comp.*, **65** (1996), 1737–1747.

[2] R. C. Baker and G. Harman, 'The Brun-Titchmarsh theorem on average', *Proc. Conf. in Honor of Heini Halberstam (Allerton Park, IL, 1995)*, Progr. Math., vol. 138, Birkhäuser, Boston, 1996, 39–103.

[3] R. C. Baker and G. Harman, 'Shifted primes without large prime factors', *Acta Arith.*, **83** (1998), 331–361.

[4] W. Banks, J. B. Friedlander, C. Pomerance and I. E. Shparlinski, 'Multiplicative structure of values of the Euler function', *High Primes and Misdemeanours: Lectures in Honour of the 60th Birthday of Hugh Cowie Williams*, Fields Institute Communications, vol. 41, Amer. Math. Soc., 2004, 29–48.

[5] P. Erdős and M. R. Murty, 'On the order of $a \pmod p$', *Proc. 5th Canadian Number Theory Association Conf.*, Amer. Math. Soc., Providence, RI, 1999, 87–97.

[6] K. Ford, 'The distribution of integers with a divisor in a given interval', *Annals Math.*, (to appear).

[7] É. Fouvry, 'Théorème de Brun–Titchmarsh; application au théorème de Fermat', *Invent. Math.*, **79** (1985), 383–407.

[8] M. Goldfeld, 'On the number of primes p for which $p + a$ has a large prime factor', *Mathematika*, **16** (1969), 23–27.

[9] G. Harman, *Prime-detecting sieves*, Princeton Univ. Press, Princeton, NJ, 2007.

[10] D. R. Heath-Brown and S. V. Konyagin, 'New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum', *Quart. J. Math.*, **51** (2000), 221–235.

[11] C. Hooley, 'On the largest prime factor of $p+a$', *Mathematika*, **20** (1973), 135–143.

[12] K.-H. Indlekofer and N. M. Timofeev, 'Divisors of shifted primes', *Publ. Math. Debrecen*, **60** (2002), 307–345.

[13] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc., Providence, RI, 2004.

[14] S. V. Konyagin and I. E. Shparlinski, *Character sums with exponential functions and their applications*, Cambridge Univ. Press, Cambridge, UK, 1999.

[15] P. Kurlberg and C. Pomerance, 'On the period of the linear congruential and power generators', *Acta Arith.*, **119** (2005), 149–169.

[16] H. L. Montgomery and R. C. Vaughan, 'The large sieve', *Mathematatika*, **20** (1973), 119–134.

[17] F. Pappalardi, 'On the order of finitely generated subgroups of $\mathbb{Q}^*$ (mod p) and divisors of $p-1$', *J. Number Theory*, **57** (1996), 207–222.

[18] C. Pomerance, 'Popular values of Euler's function', *Mathematika*, **27** (1980), 84–89.

[19] C. Pomerance and I. E. Shparlinski, 'Smooth orders and cryptographic applications', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2369** (2002), 338–348.

[20] A. Schinzel, 'A refinement of a theorem of Gerst on power residues', *Acta Arith.*, **17** (1970), 161–168.

[21] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, University Press, Cambridge, UK, 1995.