MANY SOLUTIONS TO THE S-UNIT EQUATION $a + 1 = c$

J. HA$^1$,* and K. SOUNDARARAJAN$^2$†

$^1$Department of Mathematics, Incheon National University, Incheon, Republic of Korea
e-mail: junsoo.ha.31@gmail.com

$^2$Department of Mathematics, Stanford University, Stanford, CA 94305, USA
e-mail: ksound@stanford.edu

(Received February 20, 2019; accepted March 4, 2019)

Abstract. We show that there are arbitrarily large sets $S$ of $s$ primes for which the number of solutions to $a + 1 = c$ where all prime factors of $ac$ lie in $S$ has $\gg \exp(s^{1/4}/\log s)$ solutions.

1. Introduction

Given a finite set $S$ of primes, the binary $S$-unit equation concerns solutions to $u + v = 1$ where $u$ and $v$ are $S$-units; that is, $u$ and $v$ are rational numbers whose numerator and denominator are composed only of primes in $S$. This $S$-unit equation has been extensively investigated, and we refer to [5] for a detailed overview of this equation and its generalizations. In particular, Evertse [3] has shown that the binary $S$-unit equation has at most $3 \times 7^{2s+1}$ solutions, where $s$ denotes the cardinality of the set $S$. This refines classical works of Siegel and Mahler which established the finiteness of the number of solutions to the $S$-unit equation.

While there are many naturally occurring $S$-unit equations that have very few solutions (see [4] for many examples admitting at most two solutions), it is possible to exhibit arbitrarily large sets $S$ for which the equation $u + v = 1$ has lots of solutions. In this context, Erdős, Stewart and Tijdeman [2] showed that there are arbitrarily large sets $S$ for which the equation has at least $\exp((4 - \varepsilon)(s/\log s)^{1/2})$ solutions. This was subsequently refined by Konyagin and Soundararajan [11] who showed that there are sets $S$ for
which the $S$-unit equation has at least $\exp\left(s^2 - \sqrt{2} - \varepsilon\right)$ solutions. The sets $S$ used in these constructions are special and comprise of the set of initial primes, together with a small number of primes that appear in the argument, and which are out of our control. But even for the set $S$ comprising of the first $s$ primes, it is expected that the $S$-unit equation has $\exp\left(s^{2/3} - \varepsilon\right)$ solutions, and that perhaps the general $S$-unit equation does not have more than $\exp\left(s^{2/3 + \varepsilon}\right)$ solutions (see [2] for a heuristic discussion). In the context of $S$ being the first $s$ primes (which is related to the distribution of smooth numbers), Lagarias and Soundararajan [12] showed that under the Generalized Riemann Hypothesis one has at least $\exp\left(s^{1/8} - \varepsilon\right)$ solutions, and Harper [9] has shown unconditionally that there are at least $\exp(s^6)$ solutions for some $\delta > 0$. Ha [6] has studied the analogous problem over function fields, obtaining unconditionally $\gg \exp\left(s^{1/6} - \varepsilon\right)$ solutions.

Rewrite the $S$-unit equation $u + v = 1$ as $a + b = c$ where $a$, $b$ and $c$ are coprime positive integers with all prime factors of $abc$ lying in the set $S$. In this setting, we may consider the special case when $b = 1$, where we are seeking two consecutive natural numbers $a$ and $c$ with all their prime factors lying in $S$. Konyagin and Soundararajan [11] showed that this special case too has exponentially many solutions for certain well-chosen sets $S$. Namely, they showed that there are sets $S$ for which the equation $a + 1 = c$ has at least $\exp(s^{1/16})$ solutions. This was subsequently improved by Harper [7] who showed the existence of sets $S$ for which there are at least $\exp(s^{1/6 - \varepsilon})$ solutions. In this paper we make further progress on this question, by showing that there are sets $S$ with at least $\exp(s^{1/4} / \log s)$ solutions.

**Theorem 1.** For all $s$, there exist sets $S$ of $s$ primes such that the equation

$$a + 1 = c$$

has $\gg \exp(s^{1/4} / \log s)$ solutions where all prime factors of $ac$ lie in $S$.

For the equation $a + 1 = c$, we do not know any upper bound on the number of solutions better than Evertse’s bound for the more general equation $a + b = c$. One may also ask for analogues of the results of Lagarias and Soundararajan, and Harper, where $S$ is taken to be the set of first $s$ primes. This remains unknown, but heuristic considerations (as in [12] and [2]) suggest that when $S$ is the set of first $s$ primes there are $\exp(s^{1/2 - \varepsilon})$ solutions to the equation $a + 1 = c$, and that for general sets $S$ the equation has no more than $\exp(s^{1/2 + \varepsilon})$ solutions.

**2. Deducing Theorem 1 from the main proposition**

In this section we enunciate the main technical result of the paper, from which we shall deduce Theorem 1. Let $y$ be large, and let $\ell \leq k$ be two
integer parameters. Our goal is to evaluate asymptotically

\( \mathcal{N}(y; k, \ell) = \# \{ p_1 \cdots p_k \equiv 1 \pmod{q_1 \cdots q_\ell} \} \),

where the \( p_i \) run over all primes in the interval \((y/2, y]\) and the \( q_j \) run over all primes in the interval \((y/4, y/2]\). For brevity, we write

\[ \lambda = \sum_{y/4 < q \leq y/2} \frac{1}{q} \sim \frac{\log 2}{\log y}. \]

and

\[ P = \sum_{y/2 < p \leq y} 1 \sim \frac{y}{2 \log y}. \]

We have in mind ranges where \( k \) and \( \ell \) grow with \( y \), and in the estimates below all implied constants will be absolute.

**Proposition 1.** Let \( y \geq 10 \) be a real number, and let \( \ell, k \) be integers with \( 1 \leq \ell \leq k \leq y^{1/3}/(\log y)^2 \). In the range \( \ell \leq k/2 \) we have

\[ \mathcal{N}(y; k, \ell) = \lambda^\ell P^k \left( 1 + O\left(\frac{1}{\log y}\right) \right). \]

In the range \( k/4 \leq \ell \leq k/2 \), we have

\[ \mathcal{N}(y; k, \ell) = \lambda^\ell P^k \left( 1 + O\left(\frac{1}{\log y}\right) \right) + O\left(\ell^{k-\ell}(4\lambda P)^\ell y^{k/2}\right). \]

Roughly speaking, Proposition 1 may be viewed as an average result on the equidistribution of smooth numbers in arithmetic progressions. In this sense, it is related to recent results of Harper [8] and Drappeau [1] which establish strong analogues of the Bombieri–Vinogradov theorem in this context. For our application to Theorem 1, we are essentially interested in the distribution in progressions of integers \( n \leq x \) that are \((\log x)^A\) smooth. The results of Drappeau would permit a larger level of distribution in terms of the moduli of the progressions involved, but they require a smoothness of \((\log x)^A\) for a suitably large unspecified constant \( A \), and therefore are not immediately applicable to our situation.

**Proof of Theorem 1.** Put \( \ell = \alpha k \) and \( k = y^\beta/(10 \log y) \), with \( 0 \leq \alpha \leq 1/2 \) and \( \beta \leq 1/3 - \log \log y / \log y \). With a little calculation using Proposition 1 we see that if \((1 - \alpha)(1 - \beta) \geq 1/2\) then the error term in the second assertion of the proposition is negligible compared to the main term, and we have

\[ \mathcal{N}(y; k, \ell) = \lambda^\ell P^k \left( 1 + O\left(\frac{1}{\log y}\right) \right) \geq \frac{1}{2} \lambda^\ell P^k. \]
Let $Q$ denote the set of numbers composed of exactly $\ell$ primes taken from $(y/4, y/2]$ and denote $R$ the set of numbers composed of exactly $k$ primes taken from $(y/2, y]$. We consider solutions to the congruence $r \equiv 1 \pmod{q}$ with $r \in R$ and $q \in Q$. Each solution is counted at most $k! \ell!$ times in $N(y; k, \ell)$, and therefore the number of solutions to this congruence is at least $\frac{1}{2} \lambda^\ell P^k / (k! \ell!)$. For a solution $r \equiv 1 \pmod{q}$, note that $u = (r - 1)/q$ is an integer lying below $y^k/(y/4)^k = 4^\ell y^{k-\ell}$. It follows that there is a “popular” integer $u_0$ such that the equation $r = 1 + qu_0$ has at least

\[
\frac{1}{2} \frac{\lambda^\ell P^k}{k! \ell!} 4^{\ell} y^{k-\ell} \gg \left( \frac{1}{4\ell \log y} \right)^{\ell} \left( \frac{y}{k \log y} \right)^k y^{\ell-k} \gg 10^k y^{-k\beta + (1 - \beta)\ell}
\]
solutions. If $\alpha(1 - \beta) \geq \beta$, then this number of solutions exceeds $10^k$.

The two constraints $(1 - \alpha)(1 - \beta) \geq 1/2$ and $\alpha(1 - \beta) \geq \beta$ are met by taking $\beta = 1/4$, and $\alpha = 1/3$. Take $S$ to be the set of primes in $(y/4, y]$ together with the prime factors of $u_0$. Since $u_0$ has at most $\ll (\log u_0)/\log \log u_0 \ll y^\beta$ distinct prime factors, the set $S$ has size at most $y/\log y$. Our argument above has produced

\[
\gg 10^k \gg \exp \left( \frac{y^\beta}{5 \log y} \right) \geq \exp \left( \frac{s^{1/4}}{10(\log s)^{3/4}} \right)
\]
solutions to the equation $a + 1 = c$ with all prime factors of $ac$ lying in $S$. This establishes the theorem. \hfill \Box

### 3. Proof of Proposition 1

By the orthogonality relation for Dirichlet characters, we have

\[N(y; k, \ell) = \sum_{\substack{y/4 < q_1 \leq y/2 \\ 1 \leq j \leq \ell}} 1 \varphi(q_1 \cdots q_\ell) \sum_{\chi \pmod{q_1 \cdots q_\ell}} \sum_{\substack{y/2 < p_i \leq y \\ 1 \leq i \leq k}} \chi(p_1 \cdots p_k) \]

\[= \sum_{\substack{y/4 < q_i \leq y/2 \\ 1 \leq j \leq \ell}} 1 \varphi(q_1 \cdots q_\ell) \sum_{\chi \pmod{q_1 \cdots q_\ell}} \left( \sum_{y/2 < p \leq y} \chi(p) \right)^k.
\]

We isolate the contribution of the principal character $\chi = \chi_0$ above. Since $\varphi(q_1 \cdots q_\ell) = q_1 \cdots q_\ell (1 + O(\ell/y))$, this term contributes

\[\left(1 + O\left(\frac{\ell}{y}\right)\right) \sum_{\substack{y/4 < q_i \leq y/2 \\ 1 \leq j \leq \ell}} 1 \varphi(q_1 \cdots q_\ell) \left( \sum_{y/2 < p \leq y} 1 \right)^k = \left(1 + O\left(\frac{\ell}{y}\right)\right) \lambda^\ell P^k.
\]
It remains now to estimate the contribution of the non-principal characters to (4), which is bounded by

\begin{equation}
\leq \sum_{y/4 < q_j \leq y/2} \frac{2}{q_1 \cdots q_\ell} \sum_{\chi \equiv \chi_0 \pmod{q_1 \cdots q_\ell}} \left| \sum_{y/2 < p \leq y} \chi(p) \right|^k.
\end{equation}

To estimate the contribution of the non-principal characters, we shall use the large sieve. Since the large sieve gives a bound for sums over primitive characters, we first transform (6) into a sum over primitive characters. Recall that each non-principal character \( \chi \pmod{q_1 \cdots q_\ell} \) is induced by some primitive character \( \tilde{\chi} \pmod{q_1 \cdots q_\ell} \) where \( q_1 \cdots q_\ell \) is a divisor of \( q \) for integers \( 1 \leq t \leq \ell \) define \( Q_t \) to be the set of moduli \( q \) that are composed of exactly \( t \) primes (not necessarily distinct) all taken from the interval \( (y/4, y/2] \). Thus the sum in (6) may be recast as

\begin{equation}
\sum_{t=1}^\ell \sum_{q \in Q_t} \sum_\star \left( \sum_{y/4 < q_j \leq y/2} \frac{2}{q_1 \cdots q_\ell} \right) \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k.
\end{equation}

Here the \( \star \) indicates that the sum is over primitive characters, and we used that \( \chi(p) = \tilde{\chi}(p) \) for \( y/2 < p \leq y \). Given \( q \in Q_t \) note that

\begin{equation}
\sum_{y/4 < q_j \leq y/2} \frac{2}{q_1 \cdots q_\ell} \leq \frac{2}{q} \left( \frac{\ell}{t} \right) t! \left( \sum_{y/4 < p \leq y/2} \frac{1}{p} \right) ^{\ell-t} = \frac{2}{q} \frac{\ell!}{(\ell-t)!} \lambda^{\ell-t},
\end{equation}

since we must pick \( t \) out of \( q_1, \ldots, q_\ell \) to be the \( t \) prime factors of \( q \), and these \( t \) prime factors may be permuted in at most \( t! \) ways. Since \( \ell!/ (\ell-t)! \leq \ell^t \) and \( q \geq (y/4)^t \) for \( q \in Q_t \), we conclude that the quantity in (6) may be bounded by

\begin{equation}
\ll \sum_{t=1}^\ell \left( \frac{4t}{y} \right)^t \lambda^{\ell-t} \sum_{q \in Q_t} \sum_\star \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k.
\end{equation}

We are now ready to apply the large sieve, which we now recall.

**Lemma 1.** For any sequence \( a_n \) of complex numbers, we have

\begin{equation}
\sum_{\chi \pmod{q}} \left| \sum_{n \leq N} a_n \chi(n) \right| \leq (N + q) \sum_{n \leq N} |a_n|^2,
\end{equation}

*Acta Mathematica Hungarica* 160, 2020
and

\begin{equation}
\sum_{q \leq Q} \frac{\varphi(q)}{q} \sum_{\chi \pmod{q}} \left| \sum_{n \leq N} a_n \chi(n) \right|^2 \leq (N + Q^2 - 1) \sum_{n \leq N} |a_n|^2.
\end{equation}

**Proof.** Estimate (8) follows from the orthogonality of Dirichlet characters, while (9) may be found, for example, in [10, Theorem 7.13].

From the large sieve we extract two bounds related to the quantity (7): namely,

\begin{equation}
\sum_{q \in \mathcal{Q}_t} \sum_{\tilde{\chi} \pmod{q}} \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^{2t} \ll y^t P^{2t},
\end{equation}

and

\begin{equation}
\sum_{q \in \mathcal{Q}_t} \sum_{\tilde{\chi} \pmod{q}} \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^{4t} \ll (tyP)^{2t}.
\end{equation}

Consider first the estimate (10). Write

\[ \left( \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right)^t = \sum_{n \leq y^t} a_t(n) \tilde{\chi}(n), \]

where \(a_t(n)\) denotes the number of ways of writing \(n\) as a product of \(t\) primes all from the interval \((y/2, y]\). Clearly \(a_t(n) \leq t!\) and \(\sum_n a_t(n) = P^t\). Therefore, using the large sieve estimate (8) we find

\[ \sum_{q \in \mathcal{Q}_t} \sum_{\tilde{\chi} \pmod{q}} \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^{2t} \ll |\mathcal{Q}_t| y^t \sum_{n \leq y^t} a_t(n)^2 \leq |\mathcal{Q}_t| y^t t! P^t. \]

Since \(t \leq \ell \leq y^{1/3}\) it is easy to check that \(|\mathcal{Q}_t| \leq P^t / t!\) for large \(y\), and therefore (10) follows.

The proof of (11) is similar, invoking now the large sieve estimate (9). With \(a_{2t}(n)\) defined similarly as above, (9) yields

\[ \sum_{q \in \mathcal{Q}_t} \sum_{\tilde{\chi} \pmod{q}} \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^{4t} \ll y^{2t} \sum_{n \leq y^{2t}} a_{2t}(n)^2 \leq y^{2t} (2t)! P^{2t}, \]

from which (11) follows.
If $k \geq 4t$ then from (11) and the trivial bound $\left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right| \leq P$ we get

$$
\left( \frac{4\ell}{y} \right)^t \lambda^{\ell-t} \sum_{q \in \mathcal{Q}_t} \sum_{(\text{mod } q)}^* \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k
$$

$$
\ll \left( \frac{4\ell}{y} \right)^t \lambda^{\ell-t} P^{k-4t}(tyP)^{2t} = P^k \lambda^t \left( \frac{4\ell t^2 y}{\lambda P^2} \right)^t.
$$

Since we are assuming that $\ell \leq k \leq y^{1/3}/(\log y)^2$, we may conclude that

$$
\sum_{1 \leq t \leq k/4} \left( \frac{4\ell}{y} \right)^t \lambda^{\ell-t} \sum_{q \in \mathcal{Q}_t} \sum_{(\text{mod } q)}^* \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k
$$

$$
\ll \sum_{1 \leq t \leq k/4} P^k \lambda^t (\log y)^{-t} \ll \frac{P^k \lambda^t}{\log y}.
$$

Now suppose $k/4 \leq t \leq k/2$. Interpolating between (10) and (11) using Hölder’s inequality we obtain

$$
\sum_{q \in \mathcal{Q}_t} \sum_{(\text{mod } q)}^* \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k \ll (y^t P^{2t}) \frac{4t}{2t} \left( (tyP)^{2t} \right)^{\frac{k-2t}{2t}} = t^{k-2t} P^{2t} y^{k/2}.
$$

Therefore, for $\ell \leq k/2$,

$$
\sum_{k/4 < t \leq \ell} \left( \frac{4\ell}{y} \right)^t \lambda^{\ell-t} \sum_{q \in \mathcal{Q}_t} \sum_{(\text{mod } q)}^* \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k
$$

$$
\ll \sum_{k/4 < t \leq \ell} \ell^k \lambda^t y^{k/2} \left( \frac{4P^2}{\lambda \ell y} \right)^t \ll \ell^k \lambda^t y^{k/2} \sum_{k/4 < t \leq \ell} \left( \frac{4P}{\ell} \right)^t.
$$

The right side above is dominated by the term $t = \ell$, and so we conclude that

$$
\sum_{k/4 < t \leq \ell} \left( \frac{4\ell}{y} \right)^t \lambda^{\ell-t} \sum_{q \in \mathcal{Q}_t} \sum_{(\text{mod } q)}^* \left| \sum_{y/2 < p \leq y} \tilde{\chi}(p) \right|^k \ll \ell^{k-\ell} (4\lambda P)^{\ell} y^{k/2}.
$$

The estimates (12) and (13) complete the proof of the proposition.

References

[1] S. Drappeau, Théorèmes de type Fouvry–Iwaniec pour les entiers friables, *Compos. Math.*, 151 (2015), 828–862.
[1] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society (Providence, RI, 2004).

[2] P. Erdős, C.-L. Stewart, and R. Tijdeman, Some Diophantine equations with many solutions, *Compos. Math.*, 66 (1988), 37–56.

[3] J.-H. Evertse, On equations in $S$-units and the Thue–Mahler equation, *Invent. Math.*, 75 (1984), 561–584.

[4] J.-H. Evertse, K. Győry, C.-L. Stewart, and R. Tijdeman, On $S$-unit equations in two unknowns, *Invent. Math.*, 92 (1988), 461–477.

[5] J.-H. Evertse, K. Győry, C.-L. Stewart, R. and Tijdeman, $S$-unit equations and their applications, in: *New Advances in Transcendence Theory* (Durham, 1986), Cambridge Univ. Press (Cambridge, 1988), pp. 110–174.

[6] J. Ha, Smooth polynomial solutions to a ternary additive equation, *Canad. J. Math.*, 70 (2018), 117–141.

[7] A.-J. Harper, On finding many solutions to $S$-unit equations by solving linear equations on average, arXiv:1108.3819 (2011).

[8] A.-J. Harper, Bombieri–Vinogradov and Barban–Davenport–Halberstam type theorems for smooth numbers, arXiv:1208.5992 (2012).

[9] A.-J. Harper, Minor arcs, mean values, and restriction theory for exponential sums over smooth numbers, *Compos. Math.*, 152 (2016), 1121–1158.

[10] S. Konyagin and K. Soundararajan, Two $S$-unit equations with many solutions, *J. Number Theory*, 124 (2007), 193–199.

[11] J.-C. Lagarias and K. Soundararajan, Counting smooth solutions to the equation $A + B = C$, *Proc. Lond. Math. Soc.* (3), 104 (2012), 770–798.