Acta Math. Hungar., **124 (1–2)** (2009), 115–123.
DOI: 10.1007/s10474-009-8163-5
First published online March 12, 2009

ON THE GREATEST PRIME FACTOR OF $ab+1$

K. MATOMÄKI

Department of Mathematics, 20014 University of Turku, Finland
e-mail: ksmato@utu.fi

(Received September 8, 2008; accepted November 11, 2008)

Abstract. We prove that whenever $\mathcal{A}$ and $\mathcal{B}$ are dense enough subsets of $\{1,\dots,N\}$, there exist $a\in\mathcal{A}$ and $b\in\mathcal{B}$ such that the greatest prime factor of $ab+1$ is at least $N^{1+|\mathcal{A}|/(9N)}$.

1. Introduction

Let $\mathcal{A}$ and $\mathcal{B}$ be subsets of $\{1,\dots,N\}$. Denote the sizes of $\mathcal{A}$ and $\mathcal{B}$ by A and B respectively. We investigate whether the members of the set

$$\{ab+1 \mid a\in\mathcal{A},\ b\in\mathcal{B}\}$$

have large prime factors. To this end, we write $P(n)$ for the largest prime factor of n. Sárközy and Stewart [7, Conjecture 1] have made the following conjecture.

CONJECTURE. *For each $\varepsilon>0$ there exists $N_0(\varepsilon)$ and $c(\varepsilon)$ such that if $N\geqq N_0$ and $\min\{A,B\}>\varepsilon N$, then there exists $a\in\mathcal{A}$ and $b\in\mathcal{B}$ such that*

$$P(ab+1)>c(\varepsilon)N^2.$$

Key words and phrases: greatest prime factor, Kloosterman sums, linear sieve.
2000 Mathematics Subject Classification: 11N36, 11N13.

This is also Conjecture 44 in Sárközy's collection of open problems [6]. Conjecture 45 in the same collection is the much weaker claim that, under the same assumptions, one finds $a \in \mathcal{A}$ and $b \in \mathcal{B}$ such that

$$P(ab+1) > c(\varepsilon)N^{1+c'} \tag{1}$$

for some c' independent of ε.

However, even this remains unsolved. The best result by now is the following theorem due to Stewart [8].

THEOREM. *Let $Z = \min\{A, B\}$. There are effectively computable positive numbers N_0, C_0 and c_2 such that if $N \geqq N_0$ and*

$$Z > C_0 \frac{N}{\sqrt{\log N/\log\log N}},$$

then there are $a \in \mathcal{A}$ and $b \in \mathcal{B}$ such that

$$P(ab+1) > N^{1+c_2(Z/N)^2}.$$

We will improve this by proving the following theorem.

THEOREM 1. *Let $N \geqq N_0$ and assume that*

$$A \geqq \frac{200N}{\log N} \quad \textit{and} \quad B \geqq \frac{A}{N^{A/(200N)}}.$$

Then there exist $a \in \mathcal{A}$ and $b \in \mathcal{B}$ such that

$$P(ab+1) \geqq N^{1+A/(9N)}.$$

Actually we will prove slightly more.

THEOREM 2. *Assume that*

$$A \geqq \frac{C_0 N}{\log N} \quad \textit{and} \quad B \geqq \frac{A}{N^{c_1 A/N}} \tag{2}$$

for some positive constants C_0 and c_1. Then there exist $a \in \mathcal{A}$ and $b \in \mathcal{B}$ such that

$$P(ab+1) \geqq N^{\sqrt{1+c_2 A/N}}$$

for any

$$c_2 < \frac{1 - 4c_1 - \frac{2}{C_0}}{4} \tag{3}$$

and $N \geqq N_0(c_2)$.

Theorem 1 follows immediately from Theorem 2, since

$$1+\frac{\alpha}{9} \leqq \sqrt{1+\frac{19}{81}\alpha}$$

for any $\alpha \in [0,1]$ and $c_2 = 19/81$ satisfies (3) for $C_0 = 1/c_1 = 200$.

We use Chebysev's method to prove Theorem 2. More precisely, we will evaluate the sum

$$S = \sum_{a\in\mathcal{A},\ b\in\mathcal{B}} \ \sum_{p\mid ab+1} \log p.$$

in two different ways. First directly and then splitting it up

$$S = S_1 + S_2 + S_3$$

according to the summation ranges

$$p < E, \quad p \in [E, N\eta] \quad \text{and} \quad p > N\eta,$$

where

$$E = o(N) \quad \text{and} \quad \eta = 1/(E\log^2 N)$$

will be defined later. Our improvement comes mainly from the treatment of S_3. For that sum Stewart [8] gave upper bound by first replacing both $\mathcal{A}$ and $\mathcal{B}$ by $\{1,\dots,N\}$. We use this replacement only for $\mathcal{A}$ but are able to take advantage of the thinness of the set $\mathcal{B}$. This is what lets us improve $c_2(Z/N)^2$ to $A/(9N)$ in the exponent.

Besides, we use a different argument from the previous works [7, 8] for S_1 and S_2 as well. This makes our result applicable for a wider range of A and B.

Throughout the proof we will assume that the bounds (2) hold for some positive constants C_0 and c_1 and that N is sufficiently large. Furthermore, ε will be a small positive constant, not necessarily the same at each occurrence.

2. Treatment of S, S_1 and S_2

We start with S. Let $\Lambda(n)$ be the von Mangoldt function. Then

$$S = \sum_{a\in\mathcal{A},\ b\in\mathcal{B}} \ \sum_{n\mid ab+1} \Lambda(n) - \sum_{k\geqq 2}\ \sum_{p\in\mathbb{P}} \ \sum_{\substack{a\in\mathcal{A},\ b\in\mathcal{B} \\ ab\equiv -1\ (\mathrm{mod}\ p^k)}} \log p$$

$$\geqq \sum_{a\in\mathcal{A},\ b\in\mathcal{B}} \log(ab+1) - \sum_{\substack{p^k\leqq N^2+1 \\ k\geqq 2}} B\log p\left(\frac{N}{p^k}+1\right)$$

$$\geqq \big(2 - c_1 A/N + o(1)\big) AB \log N - BN\left(1 + \sum_{\substack{p^k < N^2 \\ k \geqq 2}} \frac{\log p}{p^k}\right)$$

$$\geqq \big(2 - c_1 A/N + o(1)\big) AB \log N - \frac{AB \log N}{C_0}\left(1 + \sum_{p < 6N} \frac{\log p}{p^2 - p}\right).$$

Hence

$$S \geqq \left(2 - c_1 - \frac{2}{C_0}\right) AB \log N. \tag{4}$$

For S_1, we have

$$S_1 = \sum_{p<E} \sum_{\substack{p \mid ab+1 \\ a \in \mathcal{A},\ b \in \mathcal{B}}} \log p \leqq \sum_{b \in \mathcal{B}} \sum_{p<E} \sum_{\substack{pr \equiv 1 \, (\mathrm{mod}\, b) \\ pr \leqq Nb+1}} \log p \tag{5}$$

$$\leqq \sum_{b \in \mathcal{B}} \sum_{p<E} \log p \left(\frac{Nb+1}{bp} + 1\right) \leqq \big(1 + o(1)\big) BN \log E.$$

On the other hand, orthogonality of characters gives

$$S_2 = \sum_{a \in \mathcal{A},\ b \in \mathcal{B}} \sum_{E \leqq p \leqq N\eta} \frac{1}{\phi(p)} \sum_{\chi \, (\mathrm{mod}\, p)} \chi(ab)\overline{\chi}(-1) \log p$$

$$\leqq AB \sum_{E \leqq p \leqq N\eta} \frac{\log p}{\phi(p)} + \sum_{E \leqq p \leqq N\eta} \frac{\log p}{\phi(p)} \sideset{}{^*}\sum_{\chi \, (\mathrm{mod}\, p)} \left| \sum_{a \in \mathcal{A}} \chi(a) \sum_{b \in \mathcal{B}} \chi(b) \right|.$$

By the Cauchy–Schwarz inequality and the large sieve (see [5, Theorem 7.13]), we have

$$\sum_{q \leqq Q} \frac{q}{\phi(q)} \sideset{}{^*}\sum_{\chi \, (\mathrm{mod}\, q)} \left| \sum_{a \in \mathcal{A}} \chi(a) \sum_{b \in \mathcal{B}} \chi(b) \right| \leqq (Q^2 + N)(AB)^{1/2}.$$

Hence by partial summation

$$S_2 \leqq \big(1 + o(1)\big) AB \log N + 2N\eta (AB)^{1/2} \log N + N(AB)^{1/2} \frac{\log E}{E} \tag{6}$$

$$= \big(1 + o(1)\big) AB \log N + \frac{N(AB)^{1/2}}{E}\left(\frac{2}{\log N} + \log E\right).$$

Choosing $E = N^{c_1 A/(2N)}$, we see from (5) that

$$S_1 \leqq \big(1+o(1)\big) BN \log E \leqq \big(1+o(1)\big) \frac{c_1 AB \log N}{2} \tag{7}$$

and from (6) that

$$S_2 \leqq \big(1+o(1)\big) AB \log N + A^{1/2} B^{1/2} N^{1-c_1 A/(2N)} \left(\frac{c_1 A}{2N} \log N\right) \tag{8}$$

$$= \left(1 + o(1) + \frac{c_1 A^{1/2}}{2B^{1/2} N^{c_1 A/(2N)}}\right) AB \log N \leqq \big(1 + o(1) + c_1/2\big) AB \log N$$

by (2).

3. Treatment of S_3

Let Y be the largest prime factor of the product $\prod_{a\in\mathcal{A}} \prod_{b\in\mathcal{B}} (ab+1)$. Then

(9)

$$S_3 = \sum_{N\eta < p \leqq Y} \log p \sum_{\substack{p \mid ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B}}} 1 = \sum_{N\eta < p \leqq Y} \log p \sum_{\substack{p \mid ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B}}} \int_p^{p(1+\delta)} \frac{dy}{y \log(1+\delta)}$$

$$\leqq \int_{\frac{N\eta}{1+\delta}}^{Y} \frac{\log\big((1+\delta)y\big)}{y \log(1+\delta)} \left(\sum_{p\sim y} \sum_{\substack{p \mid ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B}}} 1 \right) dy,$$

where $p \sim P$ means $P \leqq p < (1+\delta)P$. Thus we are led to consider

$$\sum_{p\sim P} \sum_{\substack{p \mid ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B}}} 1 \leqq \sum_{b\in\mathcal{B}} \sum_{\substack{ps\equiv 1 \,(\mathrm{mod}\, b) \\ p\sim P \\ s \leqq (Nb+1)/P}} 1, \tag{10}$$

where $P \geqq N\eta/(1+\delta)$. We will apply the linear sieve to the set

$$\mathcal{F}^{(b)} = \big\{ n \sim P \mid ns \equiv 1 \,(\mathrm{mod}\, b),\ s \leqq (Nb+1)/P \big\},$$

which is counted by multiplicity.

To apply the sieve, we need information about the sets

$$\mathcal{F}_d^{(b)} = \left\{ n \in \mathcal{F}^{(b)} \mid d \mid n \right\}.$$

LEMMA 3. *Let*

$$X = \frac{\delta N \phi(b)}{b} \quad \textit{and} \quad \omega(d) = \begin{cases} 1 & \textit{if} \quad \gcd(b,d) = 1, \\ 0 & \textit{otherwise.} \end{cases}$$

Then

$$\left| \mathcal{F}_d^{(b)} \right| = \frac{\omega(d)}{d} X + O\left(b^{1/2+\varepsilon}\right).$$

PROOF. The claim is obvious if $(d,b) > 1$. Thus we can assume that $(d,b) = 1$. Then

$$\left| \mathcal{F}_d^{(b)} \right| = \sum_{\substack{dks \equiv 1 \pmod{b} \\ s \leqq (Nb+1)/P \\ dk \sim P}} 1 = \frac{1}{b} \sum_{l=0}^{b-1} \sum_{\substack{k \sim P/d \\ (k,b)=1}} e\left(\frac{l\overline{kd}}{b}\right) \sum_{s \leqq \frac{Nb+1}{P}} e\left(\frac{-ls}{b}\right)$$

$$= \frac{1}{b} \left(\frac{\delta P}{d} \cdot \frac{\phi(b)}{b} + O(b^{\varepsilon}) \right) \left(\frac{Nb+1}{P} + O(1) \right)$$

$$+ O\left(\frac{1}{b} \sum_{0 < |l| \leqq b/2} \left| \sum_{\substack{k \sim P/d \\ (k,b)=1}} e\left(\frac{l\overline{kd}}{b}\right) \right| \left| \sum_{s \leqq \frac{Nb+1}{P}} e\left(\frac{-ls}{b}\right) \right| \right).$$

By a bound for incomplete Kloosterman sums (see [3, p. 36]), we have

$$\sum_{\substack{M_1 \leqq m \leqq M_2 \\ (m,q)=1}} e\left(\frac{c\overline{m}}{q}\right) \ll q^{\frac{1}{2}+\varepsilon} (c,q)^{1/2}.$$

Hence

$$\left| \mathcal{F}_d^{(b)} \right| = \frac{\delta N}{d} \frac{\phi(b)}{b} + O\left(\frac{P}{bd} + \frac{N b^{\varepsilon}}{P} + b^{\varepsilon - 1} \right)$$

$$+ O\left(\frac{1}{b} \sum_{0 < |l| \leqq b/2} b^{1/2+\varepsilon} (b,l)^{1/2} \min\left\{ \frac{Nb+1}{P}, \frac{b}{l} \right\} \right)$$

$$= \frac{\delta N}{d} \frac{\phi(b)}{b} + O\left(b^{1/2+\varepsilon}\right) = \frac{\omega(d)}{d} X + O\left(b^{1/2+\varepsilon}\right),$$

which completes the proof. □

We write further

$$V(z) = \prod_{p<z}\left(1 - \frac{\omega(p)}{p}\right) = \prod_{p<z, p\nmid b}\left(1 - \frac{1}{p}\right)$$

$$= \frac{e^{-\gamma}}{\log z}\prod_{p\mid b}\left(1 - \frac{1}{p}\right)^{-1}\bigl(1 + o(1)\bigr) = \frac{e^{-\gamma}}{\log z}\frac{b}{\phi(b)}\bigl(1 + o(1)\bigr)$$

by Mertens' formula.

Now we are ready to apply the linear sieve (Theorem 1 of [4] with $\kappa = 1$). It gives

$$\left|\mathcal{F}^{(b)} \cap \mathbb{P}\right| \leqq F\left(\frac{\log N^{1/2-\varepsilon}}{\log P^{1/2}}\right)\frac{e^{-\gamma}\delta N}{\log P^{1/2}} + \sum_{d \leqq N^{1/2-\varepsilon}} b^{1/2+\varepsilon/2},$$

where $F(s) = 2e^{\gamma}/s$ for $0 < s \leqq 3$.

Therefore

$$\left|\mathcal{F}^{(b)} \cap \mathbb{P}\right| \leqq \frac{(4+\varepsilon)\delta N}{\log N},$$

so that by (9)

(11)

$$S_3 \leqq \sum_{b\in\mathcal{B}} \int_{\frac{N\eta}{1+\delta}}^{Y} \frac{\log y}{y\log(1+\delta)}\frac{(4+\varepsilon)\delta N}{\log N} \leqq \frac{(4+\varepsilon)BN}{\log N}\bigl(\log^2 Y - \log^2(N\eta)\bigr).$$

4. Proof of Theorem 2 and further thoughts

By (4), (7) and (8) we have

(12) $$S_3 = S - S_1 - S_2 \geqq \left(1 - 2c_1 - \frac{2}{C_0} + o(1)\right)AB\log N.$$

Together with (11) this implies that

$$\left(1 - 2c_1 - \frac{2}{C_0} + o(1)\right)AB\log N \leqq \frac{(4+\varepsilon)BN}{\log N}\bigl(\log^2 Y - \log^2(N\eta)\bigr),$$

so that

$$\log^2 Y \geqq \left(1 + \frac{1}{4}\left(1 - 4c_1 - \frac{2}{C_0} - \varepsilon\right)\frac{A}{N}\right)\log^2 N.$$

Thus for some $a \in \mathcal{A}$ and $b \in \mathcal{B}$, we have

$$P(ab+1) \geqq N^{\sqrt{1+c_2A/N}},$$

which completes the proof. □

The most critical ingredient of the proof was the treatment of S_3. As long as we have to make an estimate like (10), we cannot hope to get anything better than $P(ab+1) \geqq N^{1+cA/N}$ for some positive constant c. In order to prove something like (1) using Chebysev's method, one would need an upper bound of the type

$$\sum_{p\sim P} \sum_{\substack{p|ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B}}} 1 \leqq \frac{C\delta AB}{\log N}$$

for some positive constant C.

The left hand side here equals

$$\sum_{p\sim P} \sum_{\substack{ab\equiv -1 \pmod p \\ a\leqq N, b\leqq N}} \chi_{\mathcal{A}}(a)\chi_{\mathcal{B}}(b),$$

where $\chi_{\mathcal{F}}(n)$ is the characteristic function of the set $\mathcal{F}$. This resembles the kind of sums that Bombieri, Friedlander and Iwaniec have considered (see for example [1, Theorem 3] and a recent variant by Harman and the author [2, Lemma 2.3] avoiding a Siegel–Walfisz type condition). Unfortunately, they do not have results where the ranges of a and b are almost equal. However, if a large enough subset of either $\mathcal{A}$ or $\mathcal{B}$ factors as a product of two appropriate sets, then one would get a result like (1).

Another approach to S_3 would be to use the linear sieve as we have done. Then in order to prove (10), one would need an asymptotic formula for the sum

$$\sum_{\substack{kds=ab+1 \\ a\in\mathcal{A},\ b\in\mathcal{B} \\ s\sim S}} a_d$$

for $S \in \left[N^{1-\theta_1}, N^{1+\theta_1}\right]$ on average over $d \leqq N^{\theta_2}$ for some $\theta_1, \theta_2 > 0$.

Acknowledgements. The author thanks Glyn Harman for suggesting the problem. The author was supported by the Finnish Cultural Foundation.

References

[1] E. Bombieri, J. B. Friedlander and H. Iwaniec, Primes in arithmetic progressions to large moduli II, *Math. Ann.,* **227** (1987), 361–393.

[2] G. Harman and K. Matomäki, Some problems of analytic number theory on arithmetic semigroups, *Funct. Approx. Comment. Math.,* **38** (2008), 21–40.

[3] C. Hooley, *Applications of Sieve Methods,* volume 70 of Cambridge Tracts in Mathematics, Cambridge University Press (Cambridge, 1976).

[4] H. Iwaniec, Rosser's sieve, *Acta Arith.,* **36** (1980), 171–202.

[5] H. Iwaniec and E. Kowalski, *Analytic Number Theory,* volume 53 of American Mathematical Society Colloquium Publications (Providence, Rhode Island, 2004).

[6] A. Sárközy, Unsolved problems in number theory, *Period. Math. Hungar.,* **42** (2001), 17–35.

[7] A. Sárközy and C. L. Stewart, On prime factors of integers of the form $ab + 1$, *Publ. Math. Debrecen,* **56** (2000), 559–573.

[8] C. L. Stewart, On the greatest prime factor of integers of the form $ab + 1$, *Period. Math. Hungar.,* **43** (2001), 81–91.