JEAN-MARC DESHOUILLERS
HENRYK IWANIEC

On the greatest prime factor of $n^2 + 1$

<http://www.numdam.org/item?id=AIF_1982__32_4_1_0>
ON THE GREATEST PRIME FACTOR OF \(n^2 + 1\)

by J.-M. DESHOUILLERS and H. IWANIEC

1. Introduction.

In 1967 C. Hooley [2] (see also [3]) showed that if \(D\) is not a perfect square then the greatest prime factor of \(n^2 - D\) exceeds \(n^{11/10}\) infinitely often. In fact Hooley’s arguments yield a slightly better result with the exponent \(11/10\) replaced by any \(\theta\) less than \(\theta_0 = 1.100\,148\,3\ldots\) the solution of

\[
\frac{14}{3} (\theta - \frac{12}{11}) + \frac{28}{9} \log \left(1 + \frac{33}{14} \left(\theta - \frac{12}{11}\right)\right) - \frac{41}{33} + \frac{32}{9} \log \frac{11}{8} = 0.
\]

Among several innovative ideas in Hooley’s proof one finds a very interesting application of A. Weil’s estimate for Kloosterman sums

\[
S(n\,m;\,c) = \sum_{\substack{d \pmod{c} \atop (d, c) = 1}} e\left(n\frac{\overline{d}}{c} + m\frac{d}{c}\right) \ll (n, m, c)^{1/2} c^{3/2 + \varepsilon},
\]

where the symbol \(\overline{d}\) stands for a solution of \(d\overline{d} \equiv 1 \pmod{c}\). Recently the authors [1] investigated linear forms in Kloosterman sums \(S(nQ,m;c)\) with the variables of the summation \(n, m\) and \(c\) counted with a smooth weight function, showing (see Lemma 3) that there exists a considerable cancellation of terms.

In the paper we inject this result into the Chebyshev-Hooley method to prove the following

**Theorem.** – For any \(\varepsilon > 0\) there exist infinitely many integers \(n\) such that \(n^2 + 1\) has a prime factor greater than \(n^{8-\varepsilon}\), where \(\theta\) satisfies

\[
2 - \theta - 2 \log (2 - \theta) = \frac{5}{4} \quad (\theta = 1.202\,468\ldots).
\]
Our result can be generalized to \( n^2 - D \) by using Hooley’s arguments from [3].

The authors express their thanks to Prof. C. Hooley for interesting comments and corrections.

2. Chebyshev’s method.

Let \( x \geq 2 \) and let \( b \) be a non-negative function of \( C^\infty \)-class with support in \([x,2x]\) and the derivates of which satisfy

\[
b^{(l)}(\xi) \ll x^{-l}, \quad l = 0, 1, 2, \ldots,
\]

the implied constant in \( \ll \) depending on \( l \) alone. Denote

\[
X = \int b(\xi) \, d\xi \quad \text{and} \quad |\mathcal{A}| = \sum_{n^2 + 1 = 0 \pmod{d}} b(n).
\]

We begin with applying Chebyshev’s idea to calculate

\[
(1) \quad T(x) = \sum_p |\mathcal{A}_p| \log p = \sum_d |A_d| \wedge (d) + O(x)
\]

\[
= \sum_n b(n) \sum_{d \mid n^2 + 1} \wedge (d) + O(x) = \sum_n b(n) \log (n^2 + 1) + O(x)
\]

\[
= 2 (\log x) \int b(\xi) \, d\xi + O(x) = 2X \log x + O(x).
\]

The partial sum

\[
T_0(x) = \sum_{p \leq x} |\mathcal{A}_p| \log p
\]

\[
= \sum_{p \leq x} \sum_{v^2 + 1 = 0 \pmod{p}} (\log p) \sum_{n \equiv v \pmod{p}} b(n)
\]

can be evaluated easily by the Poisson summation formula.

**Lemma 1.** — For any \( f(\xi) \) of \( C^1 \) class with compact support in \((0, \infty)\) we have

\[
\sum_{n \equiv a \pmod{q}} f(n) = \frac{1}{q} \sum_{h} e\left(-\frac{ah}{q}\right) \hat{f}\left(\frac{h}{q}\right), \quad h \in \mathbb{Z}
\]

where \( \hat{f}(t) \) is the Fourier transform of \( f(\xi) \).
By Lemma 1
\[
\sum_{n \equiv v \pmod{p}} b(n) = \frac{1}{p} \sum_{h} e\left(-\frac{vh}{p}\right) \hat{b}\left(\frac{h}{p}\right).
\]

For \( h = 0 \) we have \( \hat{b}(0) = X \). If \( h \neq 0 \) by partial integration two times we get
\[
\hat{b}\left(\frac{h}{p}\right) = \int b(\xi) e\left(\frac{h}{p} \xi\right) d\xi
= \left(\frac{p}{2\pi i h}\right)^2 \int b''(\xi) e\left(\frac{h}{p} \xi\right) d\xi \ll h^{-2} p^2 x^{-1}.
\]

This yields
\[
\sum_{n \equiv v \pmod{p}} b(n) = \frac{X}{p} + O\left(\frac{p}{x}\right)
\]
whence
\[
(2) \quad \mathcal{T}_0(x) = X \log x + O(x).
\]

Letting \( P_x \) be the greatest prime factor of \( \prod_{x < p < 2x} (n^2 + 1) \) by (1) and (2) it follows that
\[
(3) \quad S(x) = \sum_{x < p \leq P_x} |\mathcal{A}_p| \log p = X \log x + O(x).
\]

Our aim is to estimate \( S(x) \) from above and deduce from it a lower estimate for \( P_x \).

3. Splitting up of \( S(x) \).

In what follows it will be convenient to have \( p \) counted with a smooth weight function. Therefore we arrange the sum \( S(x) \) as
\[
(4) \quad S(x) = \sum_{1 \leq j \leq J} S(x; P_j) + O(x)
\]
with \( P_j = 2^j x, \ 0 \leq j \leq J \leq 2 \log x \) and
\[
(5) \quad S(x; P_j) = \sum_{p_j < p \leq 4 P_j} |\mathcal{A}_p| C_j(p) \log p
\]
where \( C_j(\xi) \) are non-negative functions of \( C^\infty \)-class which satisfy the three following conditions

\[
\text{Supp } C_j \subseteq [P_j, 4P_j]
\]

\[
\sum_{0 < j \leq J} C_j(\xi) = \begin{cases} 1 & \text{if } 2x < \xi \leq P_x \\ O(1) & \text{if } x < \xi \leq 2x \text{ or } P_x < \xi \leq 2P_x \\ 0 & \text{otherwise} \end{cases}
\]

\( C_j^{(l)}(\xi) \ll P_j^{-l} \), with the implied constant depending on \( l \) alone. The error term \( O(x) \) in (4) comes from a trivial estimate for the contribution of primes \( p \) in the interval \((x, 2x]\) which is not completely covered.

4. Application of the sieve method.

A typical sum to be considered is

\[
S(x, P) = \sum_{P < p \leq 4P} |\mathcal{A}_p| C(p) \log p
\]

with \( x < P \leq 2P_x \). Let \( x \geq D \geq 1 \) and let \( \{\lambda_d\}_{d \leq D} \) be an upper bound sieve of level \( D \), i.e. a sequence of real numbers such that

\[
\lambda \ast 1 \geq \mu \ast 1, \quad \lambda_1 = 1, \quad \lambda_d = 0 \quad \text{for } d \geq D.
\]

We also assume that \( |\lambda_d| \leq 1 \) for all \( d \) and that \( \lambda_d = 0 \) when \( d \) is not square-free.

Thus

\[
S(x, P) \leq \sum_{d \leq D} \lambda_d \sum_{m=0 \,(\text{mod } d)} |\mathcal{A}_m| C(m) \log m.
\]

By the Poisson summation formula we write

\[
|\mathcal{A}_m| = \sum_{\nu^2 + 1 \equiv 0 \,(\text{mod } m)} \sum_{n \equiv \nu \,(\text{mod } m)} b(n) = \frac{\omega(m)}{m} X + r(\mathcal{A}_m)
\]

where \( \omega(m) \) is the number of incongruent solutions of \( \nu^2 + 1 \equiv 0 \,(\text{mod } m) \) and

\[
r(\mathcal{A}_m) = \sum_{\nu^2 + 1 \equiv 0 \,(\text{mod } m)} e\left(-\frac{\nu h}{m}\right) b\left(\frac{h}{m}\right).
\]
ON THE GREATEST PRIME FACTOR OF $n^2 + 1$

According to the above we write

(8) \[ S(x,P) \leq XV(x,P) + R(x,P) \]

where $XV(x,P)$ is considered as a main term

\[ V(x,P) = \sum_{d < D} \lambda_d \sum_{m \equiv 0 \pmod{d}} \frac{\omega(m)}{m} C(m) \log m \]

and $R(x,P)$ is the total error term

\[ R(x,P) = \sum_{d < D} \lambda_d R(x,d,P) \]

with

(9) \[ R(x,d,P) = \sum_{h \neq 0} \sum_{m \equiv 0 \pmod{d}} \frac{C(m)}{m} \log m \sum_{v^2 + 1 \equiv 0 \pmod{m}} b\left(\frac{h}{m}\right) e\left(-\frac{vh}{m}\right). \]

5. Transformation of $R(x,d,P)$.

We are searching for $D$ as large as possible for which the estimate

(10) \[ R(x,P) \ll x^{1-\varepsilon} \]

is available. By partial integration $l = [4e^{-1}]$ times we get

\[ b\left(\frac{h}{m}\right) = \left(-2\pi i \frac{h}{m}\right)^{-l} \int b^{(d)}(\xi)e\left(\frac{h}{m} \xi\right) d\xi \ll x\left(\frac{P}{|h|v}\right)^l \ll h^{-2} \]

for $|h| \geq Px^{e^{-1}} = H$, say. Hence truncating the series (9) at $h = H$ we make an error $0(\tau(d)/d)$ which contributes to $R(x,P)$ an admissible amount

\[ \sum_{d < D} \tau(d) \ll (\log D)^2 \ll (\log x)^2. \]

For the remaining terms we need an explicit formula for the solutions of

(11) \[ v^2 + 1 \equiv 0 \pmod{m}. \]

**Lemma 2. (Gauss).** — Let $m > 1$. If (11) is soluble then $m$ is represented properly as a sum of two squares

(12) \[ m = r^2 + s^2, \quad (r,s) = 1, \quad r,s > 0. \]
There is a one to one correspondence between the incongruent solutions \( v(\text{mod } m) \) of (11) and the solutions \((r,s)\) of (12) given by

\[
\frac{v}{m} = \frac{-r}{s} - \frac{r}{s(r^2 + s^2)}.
\]

**Proof.** — See [5] and [3], p. 34, eq. (68).

By Lemma 2 we get

\[
\sum_{r^2 + 1 = 0 \pmod{m}} e\left(-\frac{vh}{m}\right) = \sum_{r^2 + s^2 = m, \ r,s > 0, \ (r,s) = 1} e\left(-\frac{h \tilde{r}}{s}\right) \left\{1 + O\left(\frac{r|h|}{sm}\right)\right\}
\]

whence letting \( g(m,h) = \frac{C(m)}{m} (\log m) h \left(\frac{h}{m}\right) \) we obtain

\[
R(x,d,P) = \sum_{0 < |h| \leq H} \sum_{\substack{(r,s)=1, r, s > 0 \ \ \ \ \ r^2 + s^2 \equiv 0 \pmod{d}}} g(r^2 + s^2, h) e\left(-\frac{h \tilde{r}}{s}\right) + O\left(d^{-1} Px^{3\varepsilon - 1}\right).
\]

Here the error \(O(d^{-1} Px^{3\varepsilon - 1})\) contributes to \(R(x;P)\) less than \(Px^{3\varepsilon - 1} \log x \ll x^{1-\varepsilon}\) provided \(P \leq x^{2-5\varepsilon}\) which we henceforth assume.

For sum over \(r\) we apply Poisson's summation formula giving

\[
\sum_{\{r,s\}=1} g(r^2 + s^2, h) e\left(-\frac{h \tilde{r}}{s}\right) = \sum_{u \pmod{d}} e\left(-h \tilde{u}/s\right) \sum_{r \equiv u \pmod{d}} g(r^2 + s^2, h) = \frac{1}{ds} \sum_{u \pmod{d}} e\left(-h \tilde{u}/s - k \frac{u}{ds}\right) G(h,k;s)
\]

where \(G(h,k;s) = \int g(\xi^2 + s^2, h) e(k \xi/ds) d\xi\). Writing \(u = \alpha s + \beta d\) with \(\alpha^2 + 1 \equiv 0 \pmod{d}\) it becomes

\[
\frac{1}{ds} \sum_{\alpha^2 + 1 = 0 \pmod{d}} \sum_{k} e\left(-\frac{\alpha k}{d}\right) S(-h \tilde{d} - k; s) G(h,k;s).
\]
For $k = 0$ the Kloosterman sum $S(-hd, -k; s)$ reduces to a Ramanujan sum for which we have

$$|S(-hd, 0; s)| \leq (h, s).$$

Therefore the terms with $k = 0$ contribute less than

$$\frac{\tau(d)}{d} \sum_{0 < |d| \leq \sqrt{p}} \sum_{s \leq 2 \sqrt{p}} \frac{(h, s) \times \log p}{s} \ll \sqrt{p} \frac{x^s}{d} \ll \frac{x^{1-\varepsilon}}{d}.$$ 

Finally

$$R(x, d, p) = \frac{1}{d} \sum_{x^2 + 1 = 0 (\text{mod } d)} \sum_{0 < |d| \leq \sqrt{p}} \sum_{k \neq 0} e\left(-\frac{ak}{d}\right)$$

$$\sum_{s > 0} \frac{1}{s} S(-hd, -k; s) G(h, k; s) + O\left(\frac{x^{1-\varepsilon}}{d}\right).$$

6. Linear forms in Kloosterman sums.

Let $N, M, C \geq 1$ and $f(n, m, c)$ be a function of $C^6$ class with compact support in $[C, 2C]$ with respect to $c$ and satisfying

$$|\partial_{l_1 + l_2 + l_3}^l f(n, m, c)| \leq N^{-l_1} M^{-l_2} C^{-l_3}, \quad 0 \leq l_1, l_2, l_3 \leq 2.$$ 

In this section we borrow from [1] an estimate for the average of trilinear forms (cf. Theorem 11)

$$B_d^+(N, M, C) = \sum_{0 < n \leq N} \sum_{0 < m \leq M} \sum_{(c, d) = 1} b_{m, n} S(nd, \pm m; c) f(n, m, c)$$

where $b_{m, n}$ are arbitrary complex numbers.

**Lemma 3.** If $f(n, m, c)$ satisfies (14) then for any $\varepsilon > 0$ we have

$$\left(\sum_{D < d \leq 2D} |B_d^+(N, M, C)|\right)^2 \ll (CDM)^3 N \left(\sum_{0 < m \leq M} |b_{m, n}|^2\right)$$

$$\times \left\{\frac{D(2C^2 + MN + NC^2)(2C^2 + MN + MC^2)}{DC^2 + MN} + \leq \sqrt{D(D + M)} \cdot C^3\right\}$$

and the same upper bound holds for $(\Sigma |B_d^-|)^2$, the constant implied in $\ll$ depending on $\varepsilon$ at most.
7. Estimation of the error.

In order to make Lemma 3 applicable we first split up the sum over $s$ in $R(x,d,P)$ into $\ll \log P$ sums of the type

$$
\sum_{(s,d) = 1} a(s) s(-hd, \pm k; s) G(h, \mp k; s)
$$

where $a(s)$ is a function of $C^2$ class with support $[S,2S]$, $S \leq 2\sqrt{P}$ and satisfying $a^{(l)}(s) \ll S^{-l}$ for $l = 0,1,2$. The terms with $|k| \geq DSP^{-1/2}x^{3e} = K$, say, can be eliminated trivially: integrate by parts in $G(A, \pm k, s)$, getting

$$
G(h, \pm k, s) = \left( \frac{-ds}{2\pi ik} \right)^l \int \frac{\partial^l}{\partial \xi^l} G(\xi^2 + s^2, h) e\left( \frac{2\pi k}{ds} \right) d\xi
\ll \left( \frac{ds}{|k|\sqrt{P}} \right)^l \sqrt{P} \ll k^{-2}x^{-1}.
$$

Therefore such terms contribute to $R(x,d,P)$ less than

$$
\frac{\tau(d)}{d} \sum_{0 < |h| < H} \sum_{k \geq 1 \atop k \leq S} \frac{1}{k^2 \chi} \sum_{0 < s \leq 2\sqrt{P}} \sum 1 \ll \frac{P^{3/2}}{d \chi^2} x^\epsilon \ll \frac{x^{1-\epsilon}}{d}.
$$

For $0 \leq |h| \leq H$, $0 < |k| \leq K$ and $S < s \leq 2S$ we trivially have

$$
\frac{\partial^{l_1 + l_2 + l_3}}{\partial h^{l_1} \partial k^{l_2} \partial s^{l_3}} G(h, \mp k, s) \frac{a(s)}{s} \ll |h|^{-l_1} |k|^{-l_2} |s|^{-l_3} \frac{x^{1+12e}}{S\sqrt{P}}
$$

for $0 \leq l_1, l_2, l_3 \leq 2$. This shows that Lemma 3 is applicable with

$$
f(h,k,s) = \frac{S\sqrt{P}}{x^{1+13e}} \frac{a(s)}{s} G(h, \mp k, s)\quad \text{giving}
$$

$$
\left( \sum_{D < d \leq 2D} |R(x,d,P)| \right)^2 \ll x^{2-2e} + \frac{x^{2+4\epsilon HK}}{D^2P}
$$

$$
\times \sup_{1 \leq S \leq 2 \leq \sqrt{P}} \left\{ \frac{D^2(DS^2 + HK + HS^2)(DS^2 + HK + KS^2)}{S^2(DS^2 + HK)} + DS\sqrt{D(D+K)} \right\}
\ll x^{2-2e} + (D^2x + DP + D\chi^2) x^{48e}.
$$
Therefore (10) holds if

\begin{equation}
D \leq x^{\frac{1}{2} - 25\varepsilon}, \quad D \leq x^{2 - 10\varepsilon} P^{-1} \quad \text{and} \quad D \leq x^{1 - 10\varepsilon} P^{-\frac{1}{2}}.
\end{equation}

This result can be compared with Hooley's $D = x^{1 - \varepsilon} P^{-3/4}$.

8. Evaluation of the main term.

For $d$ square-free with $\omega(d) \neq 0$ consider

\[
L(s,d) = \sum_{m=1}^{\infty} \frac{\omega(dm)}{\omega(d)} m^{-s}.
\]

**Lemma 4.** — We have

\begin{equation}
L(s,d) = \frac{\zeta(s) L(s,\chi_d)}{\zeta(2s)} \prod_{\rho \mid d} \left(1 + \frac{1}{\rho^s}\right)^{-1}.
\end{equation}

**Proof.** — Follow the arguments of [3] on pp. 31-32 and equation (6.1). Writing

\[
C(m) \log \frac{m}{m} = \int_{(\sigma)} R(s)m^{-s} \, ds, \quad \sigma > 0
\]

by Mellin's inversion formula and partial integration two times

\[
R(s) = \int C(\xi) \frac{\log \xi}{\xi} \xi^{-1} \, d\xi \ll |s|^{-1} \log P.
\]

Therefore

\[
\sum_{m \equiv 0 \pmod{m}} \frac{\omega(m)}{m} c(m) \log m = \frac{1}{2\pi i} \int_{(\sigma)} R(s) \frac{\omega(d)}{d^s} L(s,d) \, ds
\]

\[
= R(1) \frac{\omega(d)}{d} \frac{L(1,\chi_d)}{\zeta(2)} \prod_{\rho \mid d} \left(1 + \frac{1}{\rho}\right)^{-1} + \frac{1}{2\pi i} \int_{(\frac{1}{2})} R(s) \frac{\omega(d)}{d^s} L(s,d) \, ds
\]

\[
= \frac{\omega(d)}{d} \prod_{\rho \mid d} \left(1 + \frac{1}{\rho}\right)^{-1} \frac{L(1,\chi_d)}{\zeta(2)} \int C(\xi) \frac{\log \xi}{\xi} \, d\xi + O\left(\frac{\tau^2(d)}{d P} \log P\right).
\]
This yields
\[
V(x, P) = \left( \sum_{d < D} \frac{\lambda_d}{d} \rho(d) \right) \frac{L(1, \chi_A)}{\zeta(2)} \int C(\xi) \frac{\log \frac{\xi}{\xi}}{\xi} \, d\xi + O\left( \sqrt{\frac{D}{P}} (\log x)^4 \right)
\]
where \( \rho(d) = o(d) \prod_{p|d} \left(1 + \frac{1}{p}\right)^{-1} \). Now we specify \( \lambda_d \) to be those of the Rosser sieve giving (see \cite{4})
\[
\sum_{d < D} \frac{\lambda_d}{d} \rho(d) = \prod_{p < D} \left(1 - \frac{\rho(p)}{p}\right) \left(2e^r + O\left( \frac{1}{\log D} \right) \right)
= \prod_{p < D} \left(1 - \frac{1}{p}\right) \frac{\zeta(2)}{L(1, \chi_A)} \left(2e^r + O\left( \frac{1}{\log D} \right) \right)
= \frac{2\zeta(2)}{L(1, \chi_A)} \frac{1}{\log D} \left(1 + O\left( \frac{1}{\log D} \right) \right)
\]
by the Mertens prime number theorem. Hence we conclude that
\[
V(x, P) = \frac{2}{\log D} \int C(\xi) \frac{\log \frac{\xi}{\xi}}{\xi} \, d\xi \left(1 + O\left( \frac{1}{\log D} \right) \right).
\]
We choose \( D \) equal to \( x^{1 - \sqrt{e} - \frac{1}{2}} \) thus by (6) the total main term is equal to
\[
X \sum_{0 < j < \infty} V(x, P_j) = 2(1 + O(\epsilon)) \int_x^{px} \frac{\log \frac{\xi}{\xi}}{\xi} \frac{\log (x/\sqrt{\xi})}{\sqrt{\xi}} \, d\xi
= 2(1 + O(\epsilon)) \int_1^x \frac{t \, dt}{1 - t/2} \log x
= (1 + O(\epsilon)) f(\theta) x \log x
\]
where \( f(\theta) = 4(1 - \theta - 2 \log (2 - \theta)) \) and is less than 1 for \( \theta = 1.20246887 \). The proof the Theorem follows from this and (3). One may note that the truth of Selberg’s eigenvalue conjecture leads to the lower bound \( x^{3/2 - \epsilon} \) for \( P_x \).

BIBLIOGRAPHY


Manuscrit reçu le 17 mars 1981
révisé le 15 octobre 1981.

J.-M. Deshouillers,
Université de Bordeaux I
U.E.R. de Mathématiques
et d’Informatique
Laboratoire associé au CNRS n° 226
351, cours de la Libération
F - 33405 Talence Cedex.

H. Iwaniec,
Mathematics Institute
Polish Academy of Sciences
ul. Śniadeckich 8
PL - 00950 Warszawa.