MATHEMATICS OF COMPUTATION
VOLUME 54, NUMBER 190
APRIL 1990, PAGES 895-902

FERMAT'S LAST THEOREM (CASE 1) AND THE WIEFERICH CRITERION

DON COPPERSMITH

ABSTRACT. This note continues work by the Lehmers [3], Gunderson [2], Granville and Monagan [1], and Tanner and Wagstaff [6], producing lower bounds for the prime exponent p in any counterexample to the first case of Fermat's Last Theorem. We improve the estimate of the number of residues $r \bmod p^2$ such that $r^p \equiv r \bmod p^2$, and thereby improve the lower bound on p to 7.568×10^{17}.

1. INTRODUCTION

The first case of Fermat's Last Theorem (FLTI) is the statement that, for any odd prime p, the equation $x^p + y^p = z^p$ does not have integer solutions where none of x, y, z is divisible by p. The generalized Wieferich criterion (for given q) is the statement that if FLTI fails for some prime p, then $q^p \equiv q \bmod p^2$. This criterion has been proved [1] for all $q \in \widehat{W} = \{2, 3, 5, 7, \ldots, 89\}$, the first 24 primes. It trivially holds for $q = -1$ or 0, so for convenience we write $W = \widehat{W} \cup \{-1, 0\} = \{-1, 0, 2, 3, 5, 7, \ldots, 89\}$.

The number of distinct pth powers $\pmod{p^2}$ is only p, since $(a+bp)^p \equiv a^p \pmod{p^2}$. If p violates FLTI, the generalized Wieferich criteria (for all $q \in W$) can produce a large number of distinct pth powers $(\bmod\, p^2)$, and when this number exceeds p, we establish FLTI for p.

The following lower bounds for the number of distinct pth powers $\pmod{p^2}$ have been established:

- $f_1(p, W)$, the number of integers in $[0, p^2 - 1]$, all of whose prime factors lie in W ("smooth integers");
- $f_2(p, W)$, the number of smooth integers in $[-(p^2 - 1)/2, (p^2 - 1)/2]$ [4];
- $f_3(p, W)$, the number of pairs of relatively prime smooth integers (a, b) with $-p/\sqrt{2} < a < p/\sqrt{2}$ and $1 \le b < p/\sqrt{2}$ [2].

To these we add a new bound,

- $f_4(p, W)$, the number of pairs of relatively prime smooth integers (a, b) with $b > 0$, such that $a^2 + b^2 < p^2$.

Received November 4, 1988.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 11D41; Secondary 11-04.

Clearly $f_4(p, W) \geq f_3(p, W)$.

Theorem 1. *There are at least $f_4(p, W)$ distinct pth powers $r \pmod{p^2}$, if $p \notin W$.*

Proof. Each pair (a, b) counted by $f_4(p, W)$ gives rise to a residue $r \bmod p^2$ such that $a \equiv br \bmod p^2$. Since both a and b are pth powers $\pmod{p^2}$, r is also.

Suppose two such pairs, (a_1, b_1) and (a_2, b_2), give rise to the same value of $r \pmod{p^2}$. Then from

$$a_1 \equiv b_1 r \pmod{p^2}, \qquad a_2 \equiv b_2 r \pmod{p^2}$$

we obtain

$$a_2 b_1 \equiv b_1 b_2 r \equiv a_1 b_2 \pmod{p^2},$$

whence

$$a_2 b_1 - a_1 b_2 \equiv 0 \pmod{p^2}. \tag{1}$$

As vectors in $\mathbf{R}^3$, both $(a_1, b_1, 0)$ and $(a_2, b_2, 0)$ have norm less than p, so the magnitude of their cross product, $|a_2 b_1 - a_1 b_2|$, is less than p^2. Together with (1), this implies $a_2 b_1 - a_1 b_2 = 0$. So $a_1/b_1 = a_2/b_2$ as rational numbers. Since a_j and b_j are relatively prime, both fractions are reduced to lowest terms, and $b_j > 0$ implies that both have positive denominators. Thus $(a_1, b_1) = (a_2, b_2)$.

This implies that distinct pairs (a, b) counted by $f_4(p, W)$ give rise to distinct pth power residues $r \pmod{p^2}$. □

2. Generating function

To obtain an effectively computable lower bound $\tilde{f}_4(p, W, \alpha)$ for $f_4(p, W)$, we use a generating function on two variables. We select a real number $\alpha > 1$, and an integer N such that α^{N-1} exceeds the desired bound on p, and such that our computer can handle an array with N^2 elements. We define the generating function

$$C(x, y) = \sum_{i \geq 0} \sum_{j \geq 0} c_{ij} x^i y^j$$

by

$$C(x, y) = \prod_{q \in \widehat{W}} \left(\sum_{l \geq 1} x^{\lceil \log_\alpha q^l \rceil} + \sum_{l \geq 1} y^{\lceil \log_\alpha q^l \rceil} + 1 \right). \tag{2}$$

We will compute the coefficients c_{ij} for $0 \leq i < N$, $0 \leq j < N$.

For each positive smooth integer $a = \prod_{q \in \widehat{W}} q^{(l_q)}$, define the index

$$\operatorname{ind}(a, \alpha) = \sum_{q \in \widehat{W}} \lceil \log_\alpha q^{(l_q)} \rceil.$$

Evidently, $\operatorname{ind}(a, \alpha) \geq \log_\alpha a$.

Lemma 2. *The coefficient* c_{ij} *counts pairs of relatively prime smooth positive integers* (a, b) *such that*

$$i = \operatorname{ind}(a, \alpha), \qquad j = \operatorname{ind}(b, \alpha).$$

Each pair (a, b) *is counted in only one coefficient* c_{ij}.

Proof. This follows by the properties of generating functions. In the definition of $C(x, y)$, the factor

$$\left(\sum_{l\geq 1} x^{\lceil \log_\alpha q^l \rceil} + \sum_{l\geq 1} y^{\lceil \log_\alpha q^l \rceil} + 1\right),$$

corresponding to a given $q \in \widehat{W}$, expresses the condition that q may either appear in a (to some positive power) or in b (to some power) or in neither (but not both, since a and b are relatively prime). □

Corollary 3. *We have*

$$f_3(p, W) \geq 2 \cdot \sum_{0\leq i\leq I} \sum_{0\leq j\leq J} c_{ij},$$

where $I = J = \lceil \log_\alpha(p/\sqrt{2}) \rceil - 1$.

Proof. Each pair (a, b) counted by one of the c_{ij} satisfies

$$\log_\alpha(p/\sqrt{2}) > I \geq \operatorname{ind}(a, \alpha) \geq \log_\alpha a,$$

so that $p/\sqrt{2} > a$. Similarly, $p/\sqrt{2} > b$. The pair (a, b) corresponds to two pairs counted by $f_3(p, W)$, namely (a, b) and $(-a, b)$. □

Define

$$\tilde{f}_4(p, W, \alpha) \equiv 2 \cdot \sum_{\substack{i,j \\ (\alpha^i)^2+(\alpha^j)^2<p^2}} c_{ij}.$$

Corollary 4. *For* $\alpha > 1$ *we have* $f_4(p, W) \geq \tilde{f}_4(p, W, \alpha)$.

Proof. Each pair (a, b) counted by one of the c_{ij} satisfies

$$p^2 > (\alpha^i)^2 + (\alpha^j)^2 = (\alpha^{\operatorname{ind}(a,\alpha)})^2 + (\alpha^{\operatorname{ind}(b,\alpha)})^2 \geq a^2 + b^2,$$

so that (a, b) and $(-a, b)$ are counted in $f_4(p, W)$. □

Theorem 5. *If* $\tilde{f}_4(p_0, W, \alpha) \geq p_1 > p_0$, *then FLTI holds for all* p *in the range* $p_0 \leq p < p_1$.

Proof. For fixed values α and W, $\tilde{f}_4(p, W, \alpha)$ is monotone nondecreasing in p. For p in the indicated range,

$$f_4(p, W) \geq \tilde{f}_4(p, W, \alpha) \geq \tilde{f}_4(p_0, W, \alpha) \geq p_1 > p. \quad \square$$

Procedure. Build the array of c_{ij}, using the standard techniques for computing generating functions. Starting with a known lower bound for FLTI, such as

$p_0 = 101$, repeatedly evaluate $p_k = \tilde{f}_4(p_{k-1}, W, \alpha)$, as long as $p_k > p_{k-1}$. When the process converges $(p_k = p_{k-1})$ we have found a lower bound p_k on any counterexample p to FLTI.

3. Results

We tried various values of α and got different lower bounds for the case $\widehat{W} = \{2, 3, 5, \ldots, 89\}$; these are tabulated below.

alpha	bound $(q = 89)$	size of array
1.08	6.037e17	532×532
1.05	6.608e17	841×841
1.045	6.999e17	934×934
1.041616011	7.040e17	1008×1008
1.026004485	7.568e17	1604×1604

The last two values of α correspond to the 17th and 27th roots of 2, respectively. Our bound of $p \geq 7.568 \times 10^{17}$ compares with the bound of 1.56×10^{17} obtained in [6] by estimating f_3. Only a small part of the improvement can be attributed to our use of f_4 instead of f_3. The main improvement came from our use of the generating function $C(x, y)$, whereas [6 and 2] had used an analytic approximation to f_3.

The following table compares Gunderson's results, those of Tanner and Wagstaff [6], and our results for $\alpha = 1.08$ and $\alpha = 1.05$, respectively. The first two columns are from [6]. For the last two columns we used an array of size 1024×1024.

$q(n)$	Gunderson	Tanner-Wagstaff	ours $(\alpha = 1.08)$	ours $(\alpha = 1.05)$
3	9.310e01	1.311e02	2.060e02	2.100e02
5	8.614e02	1.392e03	2.554e03	2.578e03
7	7.616e03	1.307e04	2.560e04	2.642e04
11	5.273e04	9.481e04	1.972e05	2.033e05
13	3.503e05	6.613e05	1.386e06	1.452e06
17	2.032e06	4.081e06	9.224e06	9.575e06
19	1.136e07	2.452e07	5.656e07	5.958e07
23	5.755e07	1.359e08	3.279e08	3.445e08
29	2.564e08	6.796e08	1.740e09	1.800e09
31	1.110e09	3.349e09	8.859e09	9.321e09
37	4.343e09	1.533e10	4.199e10	4.428e10
41	1.601e10	6.773e10	1.931e11	2.021e11
43	5.744e10	2.959e11	8.849e11	9.135e11
47	1.948e11	1.252e12	3.827e12	4.000e12
53	6.110e11	5.065e12	1.568e13	1.663e13
59	1.779e12	1.968e13	6.315e13	6.752e13
61	5.026e12	7.588e13	2.514e14	2.669e14
67	1.320e13	2.827e14	9.807e14	1.033e15
71	3.290e13	1.033e15	3.661e15	3.880e15

$q(n)$	Gunderson	Tanner-Wagstaff	ours ($\alpha = 1.08$)	ours ($\alpha = 1.05$)
73	7.906e13	3.755e15	1.363e16	1.456e16
79	1.762e14	1.326e16	4.992e16	5.347e16
83	3.697e14	4.610e16	1.748e17	1.908e17
89	7.145e14	1.564e17	6.037e17	6.608e17
97	1.242e15	5.150e17	2.051e18	2.286e18
101	1.985e15	1.674e18	6.954e18	7.538e18
103	2.926e15	5.419e18	2.327e19	2.535e19
107	3.835e15	1.732e19	7.534e19	8.273e19
109	4.408e15	5.516e19	2.434e20	2.736e20
113	4.107e15	1.736e20	8.045e20	8.858e20
127	2.321e15	5.248e20	2.442e21	2.734e21
131	2.686e14	1.571e21	7.593e21	[2]
137	[1]	4.640e21	2.272e22	
139		1.365e22	6.731e22	
149		3.926e22	1.967e23	
151		1.125e23	5.752e23	
157		3.188e23	1.676e24	
163		8.926e23	4.839e24	
167		2.481e24	1.344e25	
173		6.826e24	3.870e25	
179		1.858e25	1.064e26	
181		5.046e25	2.920e26	
191		1.347e26	7.929e26	
193		3.588e26	2.153e27	
197		9.502e26	5.841e27	
199		2.509e27	1.582e28	
211		6.511e27	4.236e28	
223		1.661e28	1.084e29	
227		4.218e28	2.769e29	
229		1.068e29	7.329e29	

[1] Gunderson's gives no bound for larger W.
[2] Our method ran out of storage (1024×1024) at $q = 131$ for $\alpha = 1.05$.

4. Discussion

Granularity. Our lower bound $\tilde{f}_4(p, W, \alpha)$ underestimates $f_4(p, W)$ to the extent that the logarithms are rounded up to integers in (2). That is, the integers q^l are rounded up to integral powers of α. These powers of α are sparsely distributed among the real numbers. The coarseness of the resulting approximation is analogous to granularity in a photograph.

We can lessen the effect of this granularity by choosing α closer to 1—the error approaches 0 as α approaches 1—but at the expense of increasing N, and therefore increasing the amount of storage necessary.

As an example of this effect, consider the computation of $\tilde{f}_4(p, W, \alpha)$ for

$p = 208$, $W = \{-1, 0, 2, 3\}$, and the two choices of α upon which our tables are based: 1.08 and 1.05. First let $\alpha = 1.08$ and $(a, b) = (1, 192)$. We find $\log_{1.08} 64 = 54.039$ and $\log_{1.08} 3 = 14.275$. To be conservative, the computation in (2) has rounded both logarithms up, to 55 and 15, respectively. Then the point $(a, b) = (1, 192)$ is counted in the coefficient $c_{0,70}$, which means it is being estimated as $(1, 1.08^{55+15}) \simeq (1, 218.6)$. This is too large for the bound $p = 208$: $1^2 + 218.6^2 > 208^2$. In fact, the four points $(\pm 1, 192)$ and $(\pm 192, 1)$ are discarded by this rounding procedure. For this reason we find that $\tilde{f}_4(p, W, 1.08) = 206$ underestimates $f_4(p, W) = 210$. Selecting $\alpha = 1.05$, we correctly include these four points: $\log_{1.05} 3 = 22.517$, $\log_{1.05} 64 = 85.240$, $1.05^{23+86} = 204.001$, and $1^2 + 204.001^2 < 208^2$. We find that $\tilde{f}_4(p, W, 1.05) = 210 = f_4(p, W)$.

Monotonicity. For a fixed value of α, as W grows (the Wieferich criterion is proved for more values of q), our estimate $\tilde{f}_4(p, W, \alpha)$ increases, as does $f_4(p, W)$. In the expression defining $C(x, y)$, the term 1 in the factor corresponding to a new value of q ensures that the new values of c_{ij} are at least as large as the old ones, and the other terms increase the values. (This is in contrast to the behavior of the methods in [2], where the addition of new primes to W sometimes decreased the size of the attainable bounds. This behavior is discussed in [5].) Of course, to attain these bounds, we must deal with larger arrays, and the computer storage becomes a consideration.

For a fixed array size N, to prove larger bounds for larger estimates of W, we must use larger values of α, and it is quite possible that the granularity will make it impossible to prove larger bounds after a while.

5. Improvements

If we select a value of μ such that $1 \le \mu < (4/3)^{1/4}$, and consider two disks of radius $p\mu$ and p/μ, respectively, then we can get another estimate of the number of distinct pth powers $\pmod{p^2}$. This has not given an appreciable improvement in the result.

Lemma 6. *If* $1 \le \mu < (4/3)^{1/4}$, *then the number of distinct pth powers r* $\pmod{p^2}$ *is at least* $\frac{1}{2}[f_4(p/\mu, W) + f_4(p\mu, W)]$.

Proof. We fix a pth power $r \pmod{p^2}$ and ask what points (a, b) inside either disk represent r in the sense that $a \equiv br \pmod{p^2}$, a and b are relatively prime smooth integers, and $b > 0$. We assert that $r \pmod{p^2}$ can be represented by either (1) one point in the upper half of the smaller disk, or (2) at most two points in the upper half of the annulus (the larger disk minus the smaller disk), but not both.

If we have a point (a_1, b_1) in the upper half of the smaller disk and another point (a_2, b_2) in the upper half of the larger disk, their norms are bounded by p/μ and $p\mu$, respectively, so the magnitude of their cross product is less than p^2. As before, if both points represent r, then $a_1/b_1 = a_2/b_2$ as rational numbers, whence $(a_1, b_1) = (a_2, b_2)$.

Suppose we have three points in the upper half of the large disk, P_1, P_2, P_3, all representing r. Order the points in the counter-clockwise direction, and let θ_{ij} be the angle subtended by P_i and P_j at the origin. We have either $0 \le \theta_{12} \le \pi/3$, $0 \le \theta_{23} \le \pi/3$, or $2\pi/3 \le \theta_{13} \le \pi$. So for some $i \ne j$ we have $0 \le \sin\theta_{ij} \le \sqrt{3}/2$. The magnitude of the cross product $|a_i b_j - a_j b_i|$ is bounded by

$$(p\mu)(p\mu)\sin\theta_{ij} \le p^2\mu^2\left(\sqrt{3}/2\right) < p^2 .$$

Again, since both P_i and P_j represent r, this implies that the two points are equal: $(a_i, b_i) = (a_j, b_j)$.

Thus, if we count the pairs (a, b) of relatively prime smooth integers with $b > 0$ in the smaller disk, and add half the number of such pairs in the upper half of the annulus, we will obtain a lower bound on the number of distinct pth power residues $r \pmod{p^2}$. This count is

$$f_4(p/\mu, W) + \tfrac{1}{2}[f_4(p\mu, W) - f_4(p/\mu, W)] = \tfrac{1}{2}[f_4(p/\mu, W) + f_4(p\mu, W)]. \quad \square$$

Another idea is to define an increasing sequence of positive integers γ_i and let c_{ij} count points for which $a \le \gamma_i$, $b \le \gamma_j$ (i.e., γ_i is playing the role of α^i). For example, we could have $\gamma_i = i + 1$, $0 \le i \le 40$, and subsequent values could grow as $c \cdot \alpha^i$. This would eliminate some wasted storage. Then α could become smaller (for a fixed amount of storage), and we would suffer somewhat less from the granularity of powers of α. We have not implemented this improvement.

Acknowledgments

Samuel Wagstaff introduced me to this question during his lecture at the 1988 A.M.S. Summer Conference in Brunswick, Maine, reporting on [6]. In particular, my curiosity was stimulated by his exposition of the counterintuitive behavior mentioned in §4. He also made helpful suggestions regarding the organization of the paper. An anonymous referee made further suggestions on the organization, and demanded more accuracy in my wording, for which I am thankful.

Bibliography

1. A. Granville and M. B. Monagan, *The first case of Fermat's last theorem is true for all prime exponents up to* 714,591,416,091,389, Trans. Amer. Math. Soc. **306** (1988), 329–359.
2. N. G. Gunderson, *Derivation of criteria for the first case of Fermat's last theorem and the combination of these criteria to produce a new lower bound for the exponent*, Thesis, Cornell University, 1948.
3. D. H. Lehmer and E. Lehmer, *On the first case of Fermat's last theorem*, Bull. Amer. Math. Soc. **47** (1941), 139–142.
4. B. Rosser, *On the first case of Fermat's last theorem*, Bull. Amer. Math. Soc. **45** (1939), 636–640.

5. D. Shanks and H. C. Williams, *Gunderson's function in Fermat's last theorem*, Math. Comp. **36** (1981), 291–295.

6. J. W. Tanner and S. S. Wagstaff, Jr., *New bound for the first case of Fermat's Last Theorem*, Math. Comp. **53** (1989), 743–750.

IBM Research Division, T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, New York 10598. *E-mail*: copper@ibm.com