Proc. Nat. Acad. Sci. USA
Vol. 72, No. 12, pp. 4720–4722, December 1975
Mathematics

Analysis of the subtractive algorithm for greatest common divisors

(continued fractions/partial quotients/number theory)

ANDREW C. YAO* AND DONALD E. KNUTH

Computer Science Department, Stanford University, Stanford, California 94305

Contributed by Donald E. Knuth, October 8, 1975

ABSTRACT The sum of all partial quotients in the regular continued fraction expansions of m/n, for $1 \le m \le n$, is shown to be $6\pi^{-2}\, n(\ln n)^2 + O(n \log n(\log\log n)^2)$. This result is applied to the analysis of what is perhaps the oldest nontrivial algorithm for number-theoretic computations.

An ancient Greek method (1) for finding the greatest common divisor of two positive integers by mutual subtraction (ἀνταναίρεσις) can be described as follows: "Replace the larger number by the difference of the two numbers until both are equal; then the answer is this common value." For example, the computation of gcd(18,42) requires four subtraction steps: $\{18,42\} \to \{18,24\} \to \{18,6\} \to \{12,6\} \to \{6,6\}$; the answer is 6.

Let $S(n)$ denote the average number of steps to compute $\gcd(m,n)$ by this method, when m is uniformly distributed in the range $1 \le m \le n$. We shall prove the following result:

THEOREM. $S(n) = 6\pi^{-2}(\ln n)^2 + O(\log n(\log\log n)^2)$.

1. Preliminaries

Let $\lfloor x \rfloor$ denote the largest integer less than or equal to x, and let $x \bmod y = x - y\lfloor x/y \rfloor$ be the remainder of x after division by y. We represent the continued fraction $1/(x_1 + 1/(x_2 + \cdots + 1/x_r)\cdots)$ by $/x_1,x_2,\ldots,x_r/$.

If $1 \le m \le n$, it is well known that there is a unique sequence of positive integers $q_1,\ldots,q_r$ such that $m/n = /q_1,\ldots,q_r,1/$, where $r = r(m,n) \ge 0$. The number of subtraction steps needed to compute $\gcd(m,n)$ is precisely $q_1 + \cdots + q_r$; for this is evident when m divides n, and otherwise $q_1 = \lfloor n/m \rfloor$ subtraction steps replace $\{m,n\}$ by $\{m, n \bmod m\}$, where $(n \bmod m)/m = /q_2,\ldots q_m,1/$. Therefore $S(n)$ may be interpreted as one less than the average total sum of partial quotients in the continued fraction representation of fractions with denominator n.

Let us say that (x,x',y,y') is an *H-representation* of n if

$$n = xx' + yy', \qquad x > y > 0,$$
$$\gcd(x,y) = 1, \qquad \text{and } x' \ge y' > 0. \qquad [1.1]$$

We begin our analysis with the following sharpened form of a fundamental observation due to H. A. Heilbronn (2):

LEMMA 1. *There is a 1-1 correspondence between H-representations of n and ordered pairs (m,j) where* $0 < m < 1/2\,n$ *and* $1 \le j \le r(m,n)$. *Furthermore if* (x,x',y,y') *corresponds to* (m,j), *the jth partial quotient* q_j *in the continued fraction* $m/n = /q_1,q_2,\ldots,q_r,1/$ *is* $\lfloor x/y \rfloor$.

Proof: Given $0 < m < 1/2\,n$, let $d = \gcd(m,n)$, $r = r(m,n)$, and $m/n = /q_1,q_2,\ldots,q_r,1/$. Let $m'/n = /1,q_r,\ldots,q_2,q_1/$; then $1/2\,n < m' < n$, and the correspondence $m \leftrightarrow m'$ between $(0, 1/2\,n)$ and $(1/2\,n, n)$ is 1-1.

Now let (m,r) correspond to the H-representation $(m'/d,d,(n-m')/d,d)$; and if (m,j) corresponds to (x_j,x'_j,y_j,y'_j) for some $j > 1$, let $(m,j\text{-}1)$ correspond to $(y_j,q_jx'_j+y'_j,x_j-q_jy_j,x'_j)$. It follows readily that $\lfloor x_j/y_j \rfloor = q_j$ for $1 \le j \le r$ and that $y_1 = 1$, since this construction parallels the continued fraction process for m'/n.

To complete the proof, we start with a given H-representation (x,x',y,y') and show that it corresponds to a unique (m,j). This is obvious if $x' = y'$, since the construction clearly treats every such H-representation exactly once. If $x' > y'$, let $x' = qy' + x''$ where $0 < x'' \le y'$ and $q \ge 1$. By induction on x', the H-representation $(y + qx,y',x,x'')$ corresponds uniquely to some (m,j), where $j > 1$ since $x > 1$; hence (x,x',y,y') corresponds uniquely to $(m,j\text{-}1)$. □

COROLLARY. $nS(n) = 2\Sigma\lfloor x/y \rfloor + 1 - (n \bmod 2)$, *where the sum is over all* H-*representations of* n.

Proof: By the lemma, $\Sigma\lfloor x/y \rfloor$ is the total number of subtractions to compute $\gcd(m,n)$ for $1 \le m < 1/2\,n$. It is also the total for $1/2\,n < m < n$, since $\{m,n\}$ and $\{n\text{-}m,n\}$ both reduce to $\{m,n\text{-}m\}$ after one step. Finally we add the cases $m = n$ (0 steps) and $m = 1/2\,n$ (1 step if n is even). □

2. Reduction of the problem

Let $\Sigma'\lfloor x/y \rfloor$ denote the sum over all H-representations with $x'y < 1/2\,n$. Note that

$$x/y < n/x'y = x/y + y'/x' \le x/y + 1, \qquad [2.1]$$

hence the excluded H-representations with $x'y \ge 1/2\,n$ have $\lfloor x/y \rfloor = 1$. Since $r(m,n) = O(\log n)$, we have

$$\Sigma\lfloor x/y \rfloor = \Sigma'\lfloor x/y \rfloor + O(n \log n). \qquad [2.2]$$

LEMMA 2. *Given* $x',y > 0$ *and* $x'y < 1/2\,n$, *there exist* H-*representations* (x,x',y,y') *of n if and only if*

$$\gcd(y,n) = \gcd(y,x'). \qquad [2.3]$$

And when [2.3] *holds there are exactly* $\gcd(y,n)\Pi(1\text{-}p^{-1})$ *such* H-*representations, where the product is over all primes p which divide* $\gcd(y,n)$ *but not* $y/\gcd(y,n)$.

Proof: The necessity of [2.3] is obvious, since $\gcd(x,y) = 1$. Let $d = \gcd(y,n) = \gcd(y,x') = ax' + by$. The set of all solutions (x,y') to $n = xx' + yy'$ is given by $((an + qy)/d,\ (bn - qx')/d)$, for integer q. Exactly d values of q will satisfy $0 < bn - qx' \le dx'$, i.e., $y' \le x'$; and when $y' \le x'$ we have $x = (n\text{-}yy')/x' \ge n/x' - y > y$.

It remains to count how many of these d solutions satisfy $\gcd(x,y) = 1$. If p is a prime divisor of y/d, then p does not divide an/d, hence p does not divide x. On the other hand, let $p_1,\ldots,p_r$ be the primes that divide d but not y/d; then $p_1\ldots p_r$ consecutive values of q will make $(an + qy)/d$ run through a complete residue class modulo $p_1\ldots p_r$, hence

To the memory of Hans A. Heilbronn (1908–1975).

* Present address: Dept. of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

$(p_1\text{-}1)\ldots(p_r\text{-}1)$ of these values will be relatively prime to y. □

Let $P(n)$ denote $\varphi(n)/n = \Pi(1\text{-}p^{-1})$, where the product is over all prime divisors of n, and let $P(n\backslash m)$ denote the similar product over all primes that divide n but not m. As a result of **[2.1]**, **[2.2]**, and the lemma, we have

$$\Sigma\lfloor x/y\rfloor = \sum_{d\backslash n}\ \sum_{\substack{\gcd(y,n)=d\\ 1\le y<n/2}} dP(d\backslash(y/d)) \sum_{\substack{\gcd(x',y)=d\\ 1\le x'<n/2y}} \left(\frac{n}{x'y} + O(1)\right) + O(n \log n).$$

Replacing n,y,x' respectively by md,jd,kd yields

$$\Sigma\lfloor x/y\rfloor = \sum_{m\backslash n}\ \sum_{\substack{\gcd(j,m)=1\\ j<m^2/2n}} P((n/m)\backslash j) \sum_{\substack{\gcd(k,j)=1\\ k<m^2/2nj}} \frac{m}{jk} + O(n \log n \log\log n), \quad [2.4]$$

since the excluded terms are $O(n \log n\ \sigma_{-1}(n))$, where $\sigma_{-1}(n) = \Sigma_{d\backslash n}\, 1/d = O(\log\log n)$. (See ref. 3, §22.9.)

3. Asymptotic formulas

LEMMA.

$$\sum_{p\backslash n} \frac{\log p}{p} = O(\log\log n). \quad [3.1]$$

Proof: Let n be divisible by k primes, and let c_1,c_2 be constants such that the jth prime lies between $c_1 j \log j$ and $c_2 j \log j$. Then

$$\sum_{p\backslash n} \frac{\log p}{p} \le \sum_{1\le j\le k} \frac{\log p_j}{p_j} = O\left(\sum_{1\le j\le k} \frac{\log j}{j \log j}\right) = O(\log k). \quad \square$$

Consequently

$$\sum_{d\backslash n} \frac{\mu(d)}{d} \ln\left(\frac{1}{d}\right) = \sum_{p\backslash n} \frac{\ln p}{p} P(n\backslash p) = O(\log\log n), \quad [3.2]$$

and

$$\sum_{d\backslash n} \frac{\ln d}{d} = \sum_{p^j\backslash\backslash n} \ln p\left(\frac{1}{p} + \frac{2}{p^2} + \ldots + \frac{j}{p^j}\right)\sigma_{-1}\left(\frac{n}{p^j}\right) = O((\log\log n)^2). \quad [3.3]$$

We shall now evaluate **[2.3]** step by step, beginning with the sum on k.

LEMMA.

$$\sum_{\substack{\gcd(k,j)=1\\ k<x}} \frac{1}{k} = P(j) \ln x + O(\log\log j). \quad [3.4]$$

Proof: The sum is

$$\sum_{d\backslash j} \mu(d) \sum_{kd<x} \frac{1}{kd} = \sum_{d\backslash j} \frac{\mu(d)}{d}\left(\ln \frac{x}{d} + O(1)\right). \quad \square$$

Let $\mu_m(n) = (-1)^r$ if n is the product of $r \ge 0$ distinct primes, none of which divide m, otherwise $\mu_m(n) = 0$.

LEMMA.

$$\sum_{\substack{\gcd(j,m)=1\\ j<x}} \frac{P(j\backslash d)}{j} = P(m) \ln x \sum_{\substack{\gcd(r,m)=1\\ r<x}} \frac{\mu_d(r)}{r^2} + O(\log\log m). \quad [3.5]$$

Proof: The sum is

$$\sum_{\substack{\gcd(j,m)=1\\ j<x}} \frac{1}{j} \sum_{r\backslash j} \frac{\mu_d(r)}{r} = \sum_{\substack{\gcd(r,m)=1\\ r<x}} \frac{\mu_d(r)}{r} \sum_{\substack{\gcd(j,m)=1\\ j<x/r}} \frac{1}{jr};$$

apply **[3.4]**. □

LEMMA.

$$\sum_{\substack{\gcd(j,m)=1\\ j<x}} \frac{P(j\backslash d) \ln j}{j} = \frac{1}{2} P(m)(\ln x)^2 \sum_{\substack{\gcd(r,m)=1\\ r<x}} \frac{\mu_d(r)}{r^2} + O(\log x \log\log m). \quad [3.6]$$

Proof: As in **[3.4]**, we have

$$\sum_{\substack{\gcd(k,j)=1\\ k<x}} \frac{\ln k}{k} = \sum_{d\backslash j} \mu(d) \sum_{kd<x} \frac{\ln kd}{kd}$$

$$= \sum_{\substack{d\backslash j\\ d<x}} \frac{\mu(d)}{d}\left(\frac{1}{2}\left(\ln \frac{x}{d}\right)^2 + \left(\ln \frac{x}{d}\right)(\ln d) + O(\ln d)\right)$$

$$= 1/2\ P(j)(\ln x)^2 + O(\log x \log\log j)$$

by **[3.2]**, hence the desired sum can be evaluated as in **[3.5]**. □

4. Concluding steps

Putting the results of *Section* 3 into **[2.4]**, letting N stand for $m^2/2n$, and using the fact that $P(a\backslash b)P(b) = P(ab) = P(b\backslash a)P(a)$, we have

$$\Sigma\lfloor x/y\rfloor = \sum_{m\backslash n} m \sum_{\substack{\gcd(j,m)=1\\ j<N}} \frac{P(n/m)P(j\backslash(n/m))}{j} \ln\left(\frac{N}{j}\right) + O(n\sigma_{-1}(n) \log n \log\log n)$$

$$= \sum_{m\backslash n} mP(n/m)\left(1/2\ P(m)(\ln N)^2 \sum_{\substack{\gcd(r,m)=1\\ r<N}} \frac{\mu_{n/m}(r)}{r^2}\right) + O(n\sigma_{-1}(n) \log n \log\log n)$$

$$= 1/2 \sum_{m\backslash n} mP(n/m)P(m)\left(\ln \frac{n}{2} + 2 \ln \frac{m}{n}\right)^2 \sum_{r<N} \frac{\mu_n(r)}{r^2} + O(n \log n(\log\log n)^2).$$

Since

$$\sum_{m\backslash n} m \log \frac{n}{m} = n \sum_{d\backslash n} \frac{\log d}{d} = O(n(\log\log n)^2)$$

by **[3.3]**, we can simplify this to

$$1/2 \sum_{m\backslash n} mP(n/m)P(m)(\ln n)^2 \sum_{r<N} \frac{\mu_n(r)}{r^2} + O(n \log n(\log\log n)^2).$$

We can extend the sum on r to ∞, since

$$\sum_{m\backslash n} m \sum_{r\ge N} \frac{1}{r^2} \le \sum_{\substack{m\backslash n\\ m\le\sqrt{n}}} m \sum_{r\ge 1} \frac{1}{r^2} + \sum_{\substack{m\backslash n\\ m>\sqrt{n}}} m\, O\left(\frac{n}{m^2}\right) = O\left(\sqrt{n}\sum_{m\backslash n} 1\right) = O(n^{\frac{1}{2}+\epsilon})$$

by ref. 3, §18.1. Now

$$\sum_{r\geq 1}\frac{\mu_n(r)}{r^2}=\prod_{p\backslash n}\left(1-\frac{1}{p^2}\right)=\frac{6}{\pi^2}\prod_{p\backslash n}\left(1-\frac{1}{p^2}\right)^{-1}.$$

It remains to evaluate $\Sigma_{m\backslash n} mP(n/m)P(m)$, and since this is a multiplicative function it suffices to do the evaluation when $n = p^k$; we obtain

$$\sum_{0\leq j\leq k} p^j\left(1-\frac{1}{p}\right)^2 + (p^0+p^k)\left(\left(1-\frac{1}{p}\right)-\left(1-\frac{1}{p}\right)^2\right) = p^k\left(1-\frac{1}{p^2}\right).$$

Putting everything together yields

$$\Sigma\lfloor x/y\rfloor = \frac{3}{\pi^2}n(\ln n)^2 + O(n\ \log n(\log\log n)^2),$$

and this proves the theorem in view of the corollary to the lemma of *Section 1.*

The theorem shows that the sum of all partial quotients for m/n is $O((\log n)^{2+\epsilon})$ for all but $o(n)$ values of $m \leq n$, as $n \to \infty$, and this establishes a conjecture made in ref. 4. The application in ref. 4 involves the sums of even-numbered and odd-numbered partial quotients separately. If $S_o(n)$ denotes the average of $q_1 + q_3 + q_5 + \ldots$ and $S_e(n)$ the average of $q_2 + q_4 + q_6 + \ldots$, it is easy to see from the relation between m/n and $(n-m)/n$ that $n(S_o(n) - S_e(n)) = n-1$. Hence $S_o(n) \sim S_e(n) \sim 3\pi^{-2}(\ln n)^2$.

In a sense our theorem is rather surprising, since Khintchine (5) proved that the sum of the first k partial quotients of a real number x is asymptotically $k \log_2 k$ except for x in a set of measure zero. Thus we originally expected $S(n)$ to be of order $(\log n)(\log\log n)$ instead of $(\log n)^2$.

This research was supported in part by National Science Foundation Grant DCR72-03752 A02, by the Office of Naval Research Contract NR 044-402, and by IBM Corp. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Becker, O. (1933) "Eudoxus Studien, I," *Quellen und Stud. Gesch. Math. Ast. Phys.* (B) **2**, 311–333.
2. Heilbronn, H. (1969) "On the average length of a class of finite continued fractions," in *Number Theory and Analysis*, ed. Turán, P. (Plenum Press, New York), pp. 87–96.
3. Hardy, G. H. & Wright, E. M. (1960) *An Introduction to the Theory of Numbers* (Clarendon Press, Oxford), 4th ed.
4. Knuth, D. E. (1976) "Notes on Generalized Dedekind Sums," *Acta Arith.*, in press.
5. Khintchine, A. Ya. (1935) "Metrische Kettenbruchprobleme," *Compos. Math.* **1**, 361–382.