Conspectus materiae tomi XXXIII, fasciculi 4

La revue est consacrée à la Théorie des Nombres
The journal publishes papers on the Theory of Numbers
Die Zeitschrift veröffentlicht Arbeiten aus der Zahlentheorie
Журнал посвящен теории чисел

L'adresse de la Rédaction et de l'échange	Address of the Editorial Board and of the exchange	Die Adresse der Schriftleitung und des Austausches	Адрес редакции и книгообмена

ACTA ARITHMETICA
ul. Śniadeckich 8, 00-950 Warszawa

Les auteurs sont priés d'envoyer leurs manuscrits en deux exemplaires
The authors are requested to submit papers in two copies
Die Autoren sind gebeten um Zusendung von 2 Exemplaren jeder Arbeit
Рукописи статей редакция просит предлагать в двух экземплярах

PRINTED IN POLAND

WROCŁAWSKA DRUKARNIA NAUKOWA

ACTA ARITHMETICA
XXXIII (1977)

Notes on generalized Dedekind sums*

by

DONALD E. KNUTH (Stanford, Calif.)

When Richard Dedekind prepared a commentary on one of Bernhard Riemann's fragmentary manuscripts, for publication in Riemann's *Collected Works* [13], he introduced a number-theoretic function which has recently arisen in several different contexts. Let

$$\delta(x)=\begin{cases}1, & \text{if } x \text{ is an integer,}\\ 0, & \text{otherwise;}\end{cases} \tag{0.1}$$

$$((x)) = x-\lfloor x\rfloor-\tfrac{1}{2}+\tfrac{1}{2}\delta(x) = x-\lceil x\rceil+\tfrac{1}{2}-\tfrac{1}{2}\delta(x). \tag{0.2}$$

(Here $\lfloor x\rfloor$ denotes the greatest integer $\leqslant x$ and $\lceil x\rceil$ denotes the least integer $\geqslant x$.) Then Dedekind's sum was the special case $c=0$ of the *generalized Dedekind sum*

$$\sigma(h,k,c) = 12\sum_{0\leqslant j<k}\left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{hj+c}{k}\right)\right), \tag{0.3}$$

defined for all positive integers h, k and all real values c.

Our primary purpose in this paper is to examine this sum closely, and in particular to show that $k\sigma(h,k,c)$ is always an integer which can be calculated by an efficient algorithm that deals only with integers. In view of the applications of generalized Dedekind sums, we shall also be interested in estimating and/or computing the minimum and maximum values of $\sigma(h,k,c)$ when h and k are fixed.

A secondary purpose of this paper is to illustrate the fruitful interplay between computer science and mathematics. On the one hand, we

* This research was supported in part by National Science Foundation grant GJ 36473X and by the Office of Naval Research contract NR 044-402. Some computer experiments reported herein were done via the ARPA network with the MACSYMA system, a project supported by the Advanced Research Projects Agency under ONR contract N00014-70-A-0362-0001. Reproduction in whole or in part is permitted for any purpose of the United States Government.

shall see that symbolic formula manipulation by computer is an aid to the development of number theory. The quest for efficient means of calculation is also shown to lead to nontrivial results of a purely mathematical nature that would probably not have been discovered otherwise. Furthermore, the mathematical results derived here have immediate application to the problem of generating random numbers on a computer, as discussed in [7] and [10].

In recent years important new results about generalized Dedekind sums have been derived by U. Dieter and J. Ahrens [6], and this has substantially improved the analysis of random numbers generated by a linear congruential recurrence relation. In their forthcoming book [7], they make use of the even more generalized Dedekind sum

$$s(a\,,c|x,y) = \sum_{0\leqslant j<|c|}\left(\left(\frac{j+y}{c}\right)\right)\left(\left(\frac{a(j+y)}{c}+x\right)\right), \tag{0.4}$$

where a, c are arbitrary integers and x, y are arbitrary reals. (See Rademacher and Grosswald [12] for a comprehensive survey of Dedekind sums and their generalizations.) It is not difficult to verify that

$$s(a,c|x,y) = s(a,-c|-x,y) = s(a,c|x-\lfloor x\rfloor, y-\lfloor y\rfloor) \tag{0.5}$$

and that

$$s(a,c|x,y) = \frac{1}{12}\sigma(a,c,ay+cx)+\frac{yd}{c}\left(\left(\frac{ay+cx}{d}\right)\right)-\frac{1}{2}\left(\left(\frac{ay+cx}{c}\right)\right) \tag{0.6}$$

when $c>0$, $0<y<1$ and $d=\gcd(a,c)$; hence it suffices for our purposes to work with the simpler function $\sigma(h,k,c)$.

Equations (0.5) and (0.6) follow readily from the well-known identities

$$((-x)) = -((x)), \tag{0.7}$$

$$((x+n)) = ((x)), \quad \text{integer } n, \tag{0.8}$$

$$\sum_{0\leqslant k<n}\left(\left(x+\frac{k}{n}\right)\right) = ((nx)), \quad \text{integer } n>0, \tag{0.9}$$

which are used freely below without explicit mention.

1. Preliminary transformations. For the most part we shall study $\sigma(h,k,c)$ only when h is relatively prime to k and when c is an integer. This is sufficient to establish the general behavior, because we have

Lemma 1. *Let h, k be relatively prime and let $hh'\equiv 1 \pmod{k}$. Then*

$$\sigma(dh,dk,dc) = \sigma(h,k,c), \quad \textit{integer } d>0; \tag{1.1}$$

$$\sigma(h,k,c+\theta) = \sigma(h,k,c)+6\left(\left(h'c/k\right)\right), \quad \textit{integer } c, \ \textit{real } \theta,\ 0<\theta<1. \tag{1.2}$$

Proof. For (1.1), we have

$$\sum_{0\leqslant j<dk}\left(\left(\frac{j}{dk}\right)\right)\left(\left(\frac{hj+c}{k}\right)\right) = \sum_{\substack{0\leqslant i<d\\0\leqslant j<k}}\left(\left(\frac{ik+j}{dk}\right)\right)\left(\left(\frac{hj+c}{k}\right)\right) = \sum_{0\leqslant j<k}\left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{hj+c}{k}\right)\right).$$

For (1.2) we have

$$\sum_{0\leqslant j<k}\left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{hj+c+\theta}{k}\right)\right) = \sum_{0\leqslant j<k}\left(\left(\frac{j}{k}\right)\right)\left(\left(\left(\frac{hj+c}{k}\right)\right)+\frac{\theta}{k}-\frac{1}{2}\delta\left(\frac{hj+c}{k}\right)\right) = \sum_{0\leqslant j<k}\left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{hj+c}{k}\right)\right)+0+\frac{1}{2}\left(\left(\frac{h'c}{k}\right)\right).$$

Note that this result is independent of θ. ■

Two other simple transformations will be useful:

Lemma 2. *If $0<h\leqslant k$,*

$$\sigma(nk+h,k,c) = \sigma(h,k,c), \quad \textit{integer } n\geqslant 0; \tag{1.3}$$

$$\sigma(k-h,k,c) = -\sigma(h,k,c); \tag{1.4}$$

$$\sigma(h,k,c+nk) = \sigma(h,k,c), \quad \textit{integer } n; \tag{1.5}$$

$$\sigma(h,k,-c) = \sigma(h,k,c). \tag{1.6}$$

Proof. Equations (1.3) and (1.5) are obvious; equation (1.4) follows if j is replaced by $k-j$; and equation (1.6) follows from (1.4) since $\sigma(k-h,k,c)$ obviously equals $-\sigma(h,k,-c)$. ■

The key tool in algorithms for efficient evaluation of $\sigma(h,k,c)$ is the so-called *reciprocity law* for generalized Dedekind sums, first proved in general by U. Dieter [5].

Lemma 3. *Let h, k be relatively prime and let $0\leqslant c<k$, $0<h\leqslant k$. Then*

$$\sigma(h,k,c)+\sigma(k,h,c) = f(h,k,c) \tag{1.7}$$

where

$$(1.8)\qquad f(h,k,c) = \frac{h}{k}+\frac{k}{h}+\frac{1+6\lfloor c\rfloor\lceil c\rceil}{hk} - 6\left[\frac{c}{h}\right] - 3e(h,c);$$

$$(1.9)\qquad e(h,c) = \begin{cases} 1, & \textit{if}\quad c=0 \ \textit{or}\ \ c\not\equiv 0 (\text{modulo}\, h),\\ 0, & \textit{if}\quad c>0 \ \textit{and}\ c\equiv 0(\text{modulo}\, h).\end{cases}$$

Proof. We shall defer the proof for $c=0$ until Section 5. Assume that c is an integer, $0<c<k$, and let h', k' be integers satisfying

$$(1.10)\qquad hh'+kk' = 1.$$

Since

$$(1.11)\qquad \left(\left(\frac{hj+c+1}{k}\right)\right) = \left(\left(\frac{hj+c}{k}\right)\right)+\frac{1}{k}-\frac{1}{2}\delta\left(\frac{hj+c}{k}\right)-\frac{1}{2}\delta\left(\frac{hj+c+1}{k}\right),$$

an argument like the one we used in Lemma 1 to derive (1.2) proves that

$$(1.12)\qquad \sigma(h,k,c+1) = \sigma(h,k,c)+6\left(\left(\frac{h'c}{k}\right)\right)+6\left(\left(\frac{h'(c+1)}{k}\right)\right).$$

It follows by induction on c that

$$(1.13)\qquad \sigma(h,k,c) = \sigma(h,k,0)+12\sum_{0<j<c}\left(\left(\frac{h'j}{k}\right)\right)+6\left(\left(\frac{h'c}{k}\right)\right).$$

We also have, for $0<j<k$,

$$(1.14)\qquad \left(\left(\frac{h'j}{k}\right)\right) = \left(\left(\frac{j}{hk}-\frac{k'j}{h}\right)\right) = -\left(\left(\frac{k'j}{h}-\frac{j}{hk}\right)\right) = -\left(\left(\frac{k'j}{h}\right)\right)+\frac{j}{hk}-\frac{1}{2}\delta\left(\frac{k'j}{h}\right).$$

Hence, adding (1.13) to itself with h and k interchanged,

$$\begin{aligned}\sigma(h,k,c)+\sigma(k,h,c) &= \sigma(h,k,0)+\sigma(k,h,0)+\\ &\quad +12\sum_{0<j<c}\left(\frac{j}{hk}-\frac{1}{2}\delta\left(\frac{k'j}{h}\right)\right)+6\frac{c}{hk}-3\delta\left(\frac{k'c}{h}\right)\\ &= \sigma(h,k,0)+\sigma(k,h,0)+6\frac{c^2}{hk}-6\left[\frac{c}{h}\right]+3\delta\left(\frac{c}{h}\right)\\ &= \sigma(h,k,0)+\sigma(k,h,0)+f(h,k,c)-f(h,k,0).\end{aligned}$$

When $0<\theta<1$, equations (1.2) and (1.14) imply that

$$\sigma(h,k,c+\theta)+\sigma(k,h,c+\theta) = \sigma(h,k,c)+\sigma(h,k,c)+6c/hk-3\delta(c/h).$$

Therefore (1.8) has been established for arbitrary c. ■

2. A Euclidean algorithm. The results reviewed in Section 1 lead immediately to an efficient scheme for evaluating $\sigma(h,k,c)$. Let h and k be relatively prime, with $0<h<k$, and let c be an integer with $0\leqslant c<k$. By Lemmas 2 and 3,

$$(2.1)\qquad \begin{aligned}\sigma(h,k,c) &= f(h,k,c)-\sigma(k,h,c)\\ &= f(h,k,c)-\sigma(k \bmod h,\ h,\ c \bmod h),\end{aligned}$$

hence the evaluation problem for (h,k) reduces to the same problem for $(k \bmod h, h)$. (We write "$x \bmod y$" for the remainder of x divided by y, namely $x-y\lfloor x/y\rfloor$.) The same recurrence underlies Euclid's algorithm for determining the greatest common divisor of h and k.

By writing out the process in detail, certain simplifications will become apparent. Let us set

$$(2.2)\qquad \begin{gathered} m_0 = k,\quad m_1 = h,\quad c_0 = c,\\ a_j = \lfloor m_j/m_{j+1}\rfloor,\quad b_j = \lfloor c_j/m_{j+1}\rfloor,\\ m_{j+2} = m_j \bmod m_{j+1},\quad c_{j+1} = c_j \bmod m_{j+1},\end{gathered}$$

for $0\leqslant j<t$, where t is the least integer such that

$$(2.3)\qquad m_{t+1} = 0.$$

For example, if $t=4$ we have the tableau

$$\begin{aligned} m_0 &= a_0m_1+m_2, & c_0 &= b_0m_1+c_1,\\ m_1 &= a_1m_2+m_3, & c_1 &= b_1m_2+c_2,\\ m_2 &= a_2m_3+m_4, & c_2 &= b_2m_3+c_3,\\ m_3 &= a_3m_4, & c_3 &= b_3m_4+c_4.\end{aligned}$$

Since h and k were relatively prime it follows that

$$(2.4)\qquad m_t = 1,\qquad c_t = 0.$$

Furthermore the partial quotients $a_0,\ldots,a_{t-1}$ are positive integers, and (2.2) implies that

$$(2.5)\qquad 0<m_{j+1}<m_j,\quad 0\leqslant c_j<m_j,\quad 0\leqslant b_j\leqslant a_j,\quad \text{for}\quad 0\leqslant j<t.$$

Equation (2.1) says that

$$\sigma(m_{\ +1},m_j,c_j) = f(m_{j+1},m_j,c_j)-\sigma(m_{j+2},m_{j+1},c_{j+1}),\qquad 0\leqslant j<t,$$

and $\sigma(m_{t+1}, m_t, c_t) = \sigma(0, 1, 0) = 0$; hence by iterating this recurrence we have

$$(2.6)\qquad \sigma(h, k, c) = \sum_{0\leqslant j<t} (-1)^j \left(\frac{m_{j+1}}{m_j} + \frac{m_j}{m_{j+1}} + \frac{1+6c_j^2}{m_j m_{j+1}} - 6b_j - 3e(m_{j+1}, c_j)\right).$$

This equation can be simplified in several ways. In the first place, $m_j/m_{j+1} = a_j + m_{j+2}/m_{j+1}$, so the first two terms in the summand reduce to a_j plus a telescoping series. In the second place, the $e(m_{j+1}, c_j)$ term is easy to take care of; if z is the least subscript such that $c_z = 0$, we have

$$(2.7)\qquad \sum_{1\leqslant j\leqslant t} (-1)^j e(m_{j+1}, c_j) = (t \bmod 2) + (-1)^z - \delta_{z0}.$$

In the third place, it is well known from the theory of continued fractions that $\sum_{0\leqslant j<t} (-1)^j/m_j m_{j+1}$ is a fraction whose denominator is $m_0 = k$. Therefore equation (1.11) implies by induction on c that $k\sigma(h, k, c)$ is an integer. In other words, the sum

$$(2.8)\qquad \sum_{0\leqslant j<t} (-1)^j c_j^2/m_j m_{j+1},$$

which is a rational function in the indeterminates $a_0, \ldots, a_{t-1}, b_0, \ldots, b_{t-1}$, always evaluates to a rational number with denominator m_0.

From these considerations it is almost certain that the rational function (2.8) can be simplified in general, and the author therefore used the MACSYMA symbol manipulation system [11] for $t = 5$ to guess the general form to which (2.8) simplifies. (MACSYMA is a large collection of computer programs, written to perform symbolic mathematical calculations as well as numerical operations on numbers of arbitrary precision. In particular, MACSYMA is able to simplify rational functions in any number of indeterminates. Since it took a total of less than ten minutes, from the time the author thought of simplifying (2.8) until his computer terminal typed out the simplified numerator and denominator, this can be considered a good demonstration of the use of symbolic mathematical systems in the discovery of new mathematics. In principle, of course, Euler would have been able to discover the same identity in the 18th century, if he had set himself the problem; but it would almost certainly have taken him much longer, and perhaps the lengthy formula manipulation would have been quite frustrating.) The resulting formula is stated in the following lemma.

LEMMA 4. *Let*

$$(2.9)\qquad p_0 = 1, \quad p_1 = a_0, \quad \textit{and} \quad p_j = a_{j-1}p_{j-1} + p_{j-2} \quad \textit{for} \quad 2 \leqslant j \leqslant t.$$

Then the definitions (2.2) *imply that*

$$(2.10)\qquad \sum_{0\leqslant j<t} (-1)^j \frac{c_j^2}{m_j m_{j+1}} = \frac{1}{m_0} \sum_{0\leqslant j<t} (-1)^j b_j (c_j + c_{j+1}) p_j.$$

Proof. Littlewood has said that any identity, once written down, is trivial [4]; however, we would like to understand what lies behind equation (2.10), so we don't simply wish to prove it by induction.

According to a well-known identity of Sylvester, easily proved by induction on j, we have

$$(2.11)\qquad m_0 = p_j m_j + p_{j-1} m_{j+1} \quad \text{for} \quad 0 < j < t.$$

As an alternative to induction, *this* identity can be "understood" by using Euler's characterization of the continuant polynomials p_j (see [10], exercise 4.5.3-32). Equation (2.11) leads immediately to the formula

$$(2.12)\qquad \sum_{0\leqslant j\leqslant r} (-1)^j/m_j m_{j+1} = (-1)^r p_r/m_0 m_{r+1}.$$

Now by an appropriate interchange of summation,

$$\begin{aligned} m_0 \sum_{0\leqslant j<t} (-1)^j \Big(\sum_{j\leqslant r<t} b_r m_{r+1}\Big)^2/m_j m_{j+1} &= m_0 \sum_{\substack{0\leqslant r<t\\ 0\leqslant s<t}} b_r b_s m_{r+1} m_{s+1} \sum_{0\leqslant j\leqslant \min(r,s)} (-1)^j/m_j m_{j+1} \\ &= \sum_{\substack{0\leqslant r<t\\ 0\leqslant s<t}} b_r b_s m_{r+1} m_{s+1} (-1)^{\min(r,s)} p_{\min(r,s)}/m_{\min(r,s)+1} \\ &= \sum_{0\leqslant r\leqslant s<t} b_r b_s (-1)^r p_r m_{s+1} + \sum_{0\leqslant s<r<t} b_r b_s (-1)^s p_s m_{r+1} \\ &= \sum_{0\leqslant r<t} (-1)^r b_r p_r c_r + \sum_{0\leqslant s<t} (-1)^s b_s p_s c_{s+1}. \ \blacksquare \end{aligned}$$

Using the proof technique of Lemma 4 it is possible to derive the considerably more general identity

$$(2.13)\qquad \sum_{0\leqslant j<t} (-1)^j \frac{f(c_j)}{m_j m_{j+1}} = \frac{1}{m_0} \sum_{0\leqslant j<t} (-1)^j b_j \left(\frac{f(c_j) - f(c_{j+1}) + f(0)\,\delta_{j,t-1}}{c_j - c_{j+1}}\right) p_j$$

where f is any polynomial (and hence, any function analytic at zero); this is an observation one does not expect MACSYMA to make. One of the advantages of computer-aided mathematics is that it spurs us on, as we strive to maintain our superiority over the machine.

Combining Lemma 4 with equation (2.6), and using (2.12) when $r =$ ı yields

$$(2.14)\quad \sigma(h, k, c) = \sum_{0 \leqslant j < t} (-1)^j \big(a_j - 6b_j - 3e(m_{j+1}, c_j)\big) + \\ + \frac{1}{k}\Big(h + (-1)^{t-1} p_{t-1} + 6 \sum_{0 \leqslant j < t} (-1)^j b_j (c_j + c_{j+1}) p_j\Big).$$

Therefore the following algorithm is suggested:

Algorithm 1. Let h, k be relatively prime, $0 < h < k$, and let c b an integer with $0 \leqslant c < k$. This algorithm will output the value of $\sigma(h, k, c)$ (For brevity and precision, it has been stated in Algol notation, whicl is explained below for readers not familiar with computer programmin languages.)

```
 0.  procedure sigma (integer value $h, k, c$);
 1.    begin integer $a, b, p, pp, r, s$, sigma1, sigma2;
 2.      sigma1 := 0; sigma2 := $h$;
 3.      $p := 1$; $pp := 0$; $s := 1$;
 4.      while $h > 0$ do
 5.        begin comment At this point we have $k = m_j$,  $h = m_{j+1}$
 6.          $c = c_j$,  $p = p_j$,  $pp = p_{j-1}$,  and $s = (-1)^j$ for some $j < t$
 7.          $a := \lfloor k/h \rfloor$; $b := \lfloor c/h \rfloor$; $r := c \bmod h$;
 8.            comment Now $a = a_j$, $b = b_j$, $r = c_{j+1}$;
 9.          if $r = 0$ and $c \neq 0$ then sigma1 := sigma1 $+ 3 \times s$;
10.          if $h = 1$ then sigma2 := sigma2 $+ p \times s$;
11.          sigma1 := sigma1 $+ (a - 6 \times b) \times s$;
12.          sigma2 := sigma2 $+ 6 \times b \times p \times (c + r) \times s$;
13.          $c := r$; $s := -s$;
14.          $r := k \bmod h$; $k := h$; $h := r$;
15.          $r := a \times p + pp$; $pp := p$; $p := r$;
16.        end;
17.      comment Now $s = (-1)^t$ and $p$ is the original value of $k$;
18.      if $s < 0$ then sigma1 := sigma1 $- 3$;
19.      output (sigma1 + sigma2/$p$);
20.    end.
```

(Algorithms in Algol notation are expressed as a sequence of instruc tions separated by semicolons. A sequence of instructions surrounde by **begin** and **end** acts as a single instruction, just as parentheses are use to group algebraic expressions; the instructions are performed one by on in the stated sequence. Line 0 of the program states that the followin instructions constitute a procedure for evaluating $\sigma(h, k, c)$, given th integer values h, k, and c; line 1 means that the symbols $a, b, p, pp, r,$ ı

*sigma*1, *sigma*2 are used as auxiliary integer-valued variables in the following program. If v is a variable and E is an expression, the instruction "$v := E$" means that the value of v is replaced by the present value of E. Thus, "*sigma*2 $:= h$" in line 2 means that variable *sigma*2 should be set to the (initially given) value of h, and "$s := -s$" in line 13 means that the value of variable s should be negated when we reach that point of the program. The instruction "**if** R **then** I", where R is a relation and I is an instruction, means "if R is presently true, do instruction I, otherwise do nothing." The instruction "**while** R **do** I", where R is a relation and I is an instruction, is equivalent to "**if** R **then begin** I; **while** R **do** I **end**"; in other words, the instruction I is performed zero or more times until R becomes false. Thus lines 5–16 are performed repeatedly until h is not > 0. Relationships stated between "**comment**" and the following semicolon are not part of the program, but they may be used to prove the correctness of the program; we assert that the stated relationships between the current values of the variables will hold whenever this point in the program is reached. The idea of the above program is to set *sigma*1 equal to the first sum in (2.14) and to set *sigma*2 equal to the coefficient of $1/k$; comments appearing within the program provide the basis for a rigorous proof of this fact.)

Lines 9 and 18 of the above program have the effect of subtracting 3 times (2.7) from *sigma*1. A simpler alternative would be to delete line 18 and to use the definition of $e(m_{j+1}, c_j)$ directly in line 9:

```
if $r \neq 0$ or $c = 0$ then sigma1 := sigma1 $- 3 \times s$;
```

This requires only slightly more computation and makes the algorithm slightly easier to prove, so the above sequence of instructions would be frowned upon by contemporary aesthetes of programming style. The author apologizes for his bias towards using mathematics to avoid computation.

To evaluate $\sigma(h, k, c+\theta)$ for $0 < \theta < 1$, it suffices to replace "$c+r$" by "$c+r+1$" in line 12, and to delete line 9.

The algorithm works entirely with integers, although the integers can become large when k is large. If necessary, multiples of k can be subtracted from *sigma*2 and the quotient added to *sigma*1. We have

$$b_j p_j (c_j + c_{j+1}) < a_j p_j (m_j + m_{j+1}) \\ = a_j m_0 + m_{j+1}(a_j p_j - p_{j-1}) < (a_j + 1) m_0$$

by (2.11). Therefore the size of numbers in line 12 is reasonably well controlled. ■

3. Extreme values of Dedekind sums. Let h, k be relatively prime, $0 < h < k$. We shall now develop an algorithm to calculate an integer c

which maximizes $\sigma(h, k, c)$, for fixed h and k. Such an algorithm can also be used to find the c which minimizes $\sigma(h, k, c)$, since c minimizes $\sigma(h, k, c)$ if and only if it maximizes $\sigma(k-h, k, c)$, by (1.4).

The alternating character of our formulas in Section 2, i.e., the presence of the factor $(-1)^j$, makes it very difficult to see how to maximize $\sigma(h, k, c)$; indeed, the form of the answer we shall obtain shows that it would be very difficult to discover the correct value of c by working directly with the Euclidean construction of Section 2. For our present purposes it is much more convenient to work with a "subtractive" process, using $\lceil x\rceil$ in place of $\lfloor x\rfloor$ in the previous formulas.

Let us set

$$\begin{gathered} M_0 = k, \quad M_1 = h, \quad C_0 = c, \\ A_j = \lceil M_j/M_{j+1}\rceil, \quad B_j = \lceil C_j/M_{j+1}\rceil, \\ M_{j+2} \equiv (-M_j)\bmod M_{j+1}, \quad C_{j+1} = (-C_j)\bmod M_{j+1}, \end{gathered} \tag{3.1}$$

for $0 \leqslant j < T$, where T is the least integer such that

$$M_{T+1} = 0. \tag{3.2}$$

(Cf. (2.2), (2.3).) For example, if $T = 4$ we have the tableau

$$\begin{aligned} M_0 &= A_0M_1 - M_2, & C_0 &= B_0M_1 - C_1, \\ M_1 &= A_1M_2 - M_3, & C_1 &= B_1M_2 - C_2, \\ M_2 &= A_2M_3 - M_4, & C_2 &= B_2M_3 - C_3, \\ M_3 &= A_3M_4; & C_3 &= B_3M_4 - C_4. \end{aligned}$$

As in the additive process we have

$$M_T = 1, \qquad C_T = 0. \tag{3.3}$$

The analog of (2.5) is

$$0 < M_{j+1} < M_j, \quad 0 \leqslant C_j < M_j, \quad 0 \leqslant B_j \leqslant A_j, \quad A_j \geqslant 2, \quad \text{for} \quad 0 \leqslant j < T. \tag{3.4}$$

The important advantage of the subtractive process is that we now have an additive recurrence,

$$\begin{aligned} \sigma(M_{j+1}, M_j, C_j) &= f(M_{j+1}, M_j, C_j) - \sigma(M_j, M_{j+1}, C_j) \\ &= f(M_{j+1}, M_j, C_j) + \sigma(M_{j+2}, M_{j+1}, C_j), \quad 0 \leqslant j < T, \end{aligned}$$

by equations (1.3), (1.4); hence the $(-1)^j$ factor does not appear in

$$\sigma(h, k, c) = \sum_{0 \leqslant j < T}\left(\frac{M_{j+1}}{M_j} + \frac{M_j}{M_{j+1}} + \frac{1+6C_j^2}{M_jM_{j+1}} - 6B_j + 3E(M_{j+1}, C_j)\right), \tag{3.5}$$

$$E(M, C) = \begin{cases} -1, & \text{if} \quad C = 0; \\ 0, & \text{if} \quad C \neq 0 \text{ and } C \bmod M = 0; \\ +1, & \text{if} \quad C \bmod M \neq 0. \end{cases} \tag{3.6}$$

Our interest in (3.5) rests solely in those terms which depend on c; we obtain the maximum of $\sigma(h, k, c)$ if and only if we maximize

$$\sum_{0 \leqslant j < T}\left(\frac{C_j^2}{M_jM_{j+1}} - B_j + \frac{1}{2}E(M_{j+1}, C_j)\right) \tag{3.7}$$

over all the appropriate choices of $B_0, B_1, \ldots, B_{T-1}$.

THEOREM 1. *Let h, k be relatively prime integers, with $0 < h < k$. The maximum value of $\sigma(h, k, c)$, over all integers c in the range $0 < c < k$, occurs when $B_j = 1$ for $0 \leqslant j < T$ in the subtractive process* (3.1), (3.2).

Proof. Let

$$P_0 = 1, \quad P_1 = A_0, \quad P_j = A_{j-1}P_{j-1} - P_{j-2} \tag{3.8}$$

be the subtractive analog of (2.9). Then it is easy to verify that the analogs of (2.11), (2.12) are

$$M_0 = P_jM_j - P_{j-1}M_{j+1}, \quad 0 < j \leqslant T; \tag{3.9}$$

$$\sum_{0 \leqslant j \leqslant r} 1/M_jM_{j+1} = P_r/M_0M_{r+1}. \tag{3.10}$$

Since each $A_j \geqslant 2$, we have $P_j \geqslant 2P_{j-1} - P_{j-2}$; i.e., the P's are convex,

$$P_j - P_{j-1} \geqslant P_{j-1} - P_{j-2}. \tag{3.11}$$

It follows that

$$P_j \geqslant 2P_{j-1} - 2P_{j-2} + 2P_{j-3} - \ldots, \tag{3.12}$$

where we may assume that $P_{-1} = P_{-2} = \ldots = 0$. Equality holds in (3.12) iff j is odd and $A_{j-1} = A_{j-3} = \ldots = 2$. A similar inequality applies to the M's, i.e.,

$$M_j \geqslant 2M_{j+1} - 2M_{j+2} + 2M_{j+3} - \ldots \tag{3.13}$$

Now let c be a value which maximizes $\sigma(h, k, c)$. We may assume that $c \leqslant \frac{1}{2}k$, by (1.6); and under this assumption we shall prove that $B_0 = B_1 = \ldots = B_{T-1} = 1$ yields the maximum.

For convenience in notation, suppose we have proved that $B_0 = B_1 = B_2 = 1$ and we wish to show that $B_3 = 1$; essentially the same argument will work for all B_j. If $C_3 \neq 0$, the first three terms of (3.7) are

$$\begin{aligned} \varphi(C_3) &= \frac{(M_1-M_2+M_3-C_3)^2}{M_0M_1} + \frac{(M_2-M_3+C_3)^2}{M_1M_2} + \frac{(M_3-C_3)^2}{M_2M_3} - \frac{3}{2} \\ &= \frac{C_3^2P_2}{M_0M_3} - 2C_3\left(\frac{M_1-M_2+M_3}{M_0M_1} + \frac{-M_2+M_3}{M_1M_2} + \frac{M_3}{M_2M_3}\right) + W \\ &= \frac{1}{M_0}\left(\frac{C_3^2P_2}{M_3} - 2C_3(P_2-P_1+P_0)\right) + W \end{aligned} \tag{3.14}$$

where W is independent of C_3. Since $P_2/M_3M_0 > 0$, the minimum of this quadratic $\varphi(C_3)$ occurs when the derivative is zero, i.e., when

$$C_3/M_3 = (P_2-P_1+P_0)/P_2;$$

and $(P_2-P_1+P_0)/P_2 \geqslant 1/2$, by (3.12).

We may now conclude that $C_3 \leqslant \frac{1}{2}M_3$, by using the following argument. Suppose $C_3 > \frac{1}{2}M_3$, and let c' be the value defined by

$$c' = C_0' = M_1 - C_1', \quad C_1' = M_2 - C_2', \quad C_2' = M_3 - C_3', \quad C_3' = M_3 - C_3.$$

Since the minimum of $\varphi(C_3)$ occurs at a point $\geqslant \frac{1}{2}M_3$, we have $\varphi(C_3') \geqslant \varphi(C_3)$. Furthermore $\sigma(M_4, M_3, C_3) = \sigma(M_4, M_3, M_3-C_3)$ by (1.6), so we have $\sigma(h, k, c') \geqslant \sigma(h, k, c)$. The optimality of c implies that $\sigma(h, k, c') = \sigma(h, k, c)$, hence $\varphi(C_3') = \varphi(C_3)$, hence $(P_0-P_1+P_2)/P_2 = 1/2$, but this is impossible.

A different argument is used to show that $C_4 \leqslant \frac{1}{2}M_4$, since $(P_3-P_2+ +P_1-P_0)/P_3 = 1/2$ is possible when $A_2 = A_0 = 2$. However, $C_4 > \frac{1}{2}M_4$ implies that $C_2 > M_3 - M_4 + \frac{1}{2}M_4 = \frac{1}{2}M_2$, a contradiction.

The fourth term of (3.7) is

$$\psi(B_3) = \frac{(B_3M_4-C_4)^2}{M_3M_4} - B_3 + \frac{1}{2}E(M_4, C_3),$$

and thes quadratic $\psi(B_3)$ has its minimum when

$$\frac{2(B_3M_4-C_4)}{M_3} - 1 = 0,$$

i.e., when $C_3 = \frac{1}{2}M_3$. Therefore if $B_3 > 1$, decreasing B_3 (while holding C_4 fixed) causes both $\psi(B_3)$ and $\varphi(C_3)$ to increase. It follows that $B_3 = 1$ when $C_3 \neq 0$.

All of our arguments so far have been made under the assumption that the C_j were nonzero. We have proved that there is a subscript $z \geqslant 1$ such that $B_j = 1$ for $0 \leqslant j < z$, and $B_j = 0$ for $z \leqslant j < T$. It remains to choose the best value of z. Formula (3.7) reduces to

$$\sum_{0\leqslant j<z} \frac{\Big(\sum\limits_{j<r\leqslant z}(-1)^r M_r\Big)^2}{M_jM_{j+1}} - \frac{T+1}{2}. \tag{3.15}$$

For example, the value of (3.7) when $z = 4$ is

$$\frac{(M_1-M_2+M_3-M_4)^2}{M_0M_1} + \frac{(M_2-M_3+M_4)^2}{M_1M_2} + \frac{(M_3-M_4)^2}{M_2M_3} + \\ + \frac{M_4^2}{M_3M_4} - \frac{T+1}{2}.$$

Let

$$\varphi_4(X) = \frac{(M_1-M_2+M_3-X)^2}{M_0M_1} + \frac{(M_2-M_3+X)^2}{M_1M_2} + \frac{(M_3-X)^2}{M_2M_3} + \frac{X^2}{M_3M_4}.$$

Then

$$\varphi_4(M_4) - \varphi_3(M_3) = \varphi_4(M_4) - \varphi_4(0) = \frac{1}{M_0}\left(M_4^2\frac{P_3}{M_4} - 2M_4(P_2-P_1+P_0)\right) \\ = \frac{M_4}{M_0}(P_3-2P_2+2P_1-2P_0) \geqslant 0.$$

Similar arguments apply for all z, hence

$$\varphi_1(M_1) \leqslant \varphi_2(M_2) \leqslant \ldots \leqslant \varphi_T(M_T). \ \blacksquare$$

Note that Theorem 1 only finds the maximum over the range $0 < c < k$; it is possible that an even larger value will occur when $c = 0$. In fact, this happens if and only if

$$\varphi_T(M_T) < \tfrac{1}{2} \tag{3.16}$$

in the notation of the above proof. If $A_j = 2$ for any j, we have

$$\varphi_T(M_T) \geqslant \varphi_j(M_j) \geqslant M_j^2/M_{j-1}M_j = M_j/(2M_j - M_{j+1}) \geqslant 1/2,$$

so $\sigma(h, k, 0)$ will not be maximum. But on the other hand if $A_j = x$ for all j, we have $\varphi_T(M_T) \sim T/x$ as $x\to\infty$, hence the maximum will occur at $c = 0$ for sufficiently large x.

The proof of Theorem 1 demonstrates that there is exactly one value of $0 < c \leqslant \frac{1}{2}k$ where the maximum occurs. (For if $\varphi_T(M_T) = \varphi_{T-1}(M_{T-1})$, we have T even and $A_{T-2} = A_{T-4} = \ldots = A_0 = 2$. But then $B_{T-1} = 0$ implies that $C_{T-2} = M_{T-1} > \frac{1}{2}M_{T-2}$.)

It is possible to generalize the proof of Theorem 1 in order to find the maximum of $\sigma(h, k, c)$ over all real c; it turns out that the maximum, over all real c including $c = 0$, occurs when

$$c = M_1 - M_2 + \ldots + (-1)^T M_{T-1} + \tfrac{1}{2}(-1)^{T+1}. \tag{3.17}$$

Let us now connect up the additive and subtractive processes. In order to simplify the formulas we will be obtaining, we shall assume that t is always even in the additive Euclidean algorithm. (If t is odd, replace

$$\begin{array}{lcl} m_{t-1} = a_{t-1}m_t & \text{by} & m_{t-1} = (a_{t-1}-1)m_t + m_{t+1} \\ m_t = 1 & & m_t = (1)m_{t+1} \\ m_{t+1} = 0 & & m_{t+1} = 1 \\ & & m_{t+2} = 0 \end{array} \tag{3.18}$$

and increase t by 1.) This yields an even number of partial quotients $a_0, a_1, \ldots, a_{t-1}$ which we shall call the *canonical sequence* for (h, k). The formulas that we have derived for the evaluation of $\sigma(h, k, c)$ still hold, for the canonical sequence, since Lemma 3 includes the case $h = k = 1$.

The subtractive quotients $A_0, \ldots, A_{T-1}$ can be expressed readily in terms of the canonical sequence, as

$$a_0+1,\ (a_1-1)\times 2,\ a_2+2,\ (a_3-1)\times 2,\ \ldots,\ a_{t-2}+2,\ (a_{t-1}-1)\times 2, \tag{3.19}$$

where $(a_j-1)\times 2$ stands for a sequence of a_j-1 elements each equal to 2. Thus in particular

$$T = \sum_{\substack{0\leqslant j<t\\ j\ \text{odd}}} a_j. \tag{3.20}$$

For example, if $h = 3141592621$ and $k = 2^{35} = 34359738368$, the additive partial quotients are

$$10, 1, 14, 1, 7, 1, 1, 1, 3, 3, 3, 5, 2, 1, 8, 7, 1, 4, 1, 2, 4, 2,$$

and the subtractive ones are

$$11, 16, 9, 3, 5, 2, 2, 5, 2, 2, 2, 2, 4, 10, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 6, 2.$$

The canonical sequence for the pair of numbers whose subtractive quotients are $A_0, A_1, \ldots, A_{T-1}$, when all $A_j \geqslant 3$, is

$$A_0-1, 1, A_1-2, 1, \ldots, A_{T-1}-2, 1.$$

If $h = k-1$ the additive quotients are $1, k-1$ and the subtractive ones are $(k-1)\times 2$. Thus the subtractive process can be exponentially slower than the additive one, although it can also be twice as fast in favorable cases.

The subtractive convergents $M_0, M_1, \ldots, M_T$ are easily expressed as

$$m_0, \langle jm_2+m_3\rangle_{a_1}^{1}, \langle jm_4+m_5\rangle_{a_3}^{1}, \ldots, \langle jm_t+m_{t+1}\rangle_{a_{t-1}}^{1}, \tag{3.21}$$

where $\langle f(j)\rangle_a^1$ stands for $f(a), f(a-1), \ldots, f(1)$.

It is now possible to express the number c of Theorem 1 in terms of the canonical partial quotients, so that we obtain an efficient algorithm for the evaluation of c. The first $a = a_1$ steps of the subtractive process give

$$\begin{aligned}
C_0 &= M_1 - C_1, &\quad \text{i.e.,}\quad C_0 &= am_2+m_3-C_1,\\
C_1 &= M_2 - C_2, & C_1 &= (a-1)m_2+m_3-C_2,\\
&\cdots\cdots\cdots & &\cdots\cdots\cdots\cdots\cdots\cdots\\
C_{a-1} &= M_a - C_a, & C_{a-1} &= m_2+m_3-C_a,
\end{aligned}$$

and it follows that

$$C_0 = \begin{cases} (a/2)m_2 + C_a, & a \text{ even};\\ \big((a+1)/2\big)m_2+m_3-C_a, & a \text{ odd}. \end{cases} \tag{3.22}$$

Similarly, if $b = a_3$ we have

$$C_a = \begin{cases} (b/2)m_4 + C_{a+b}, & b \text{ even};\\ \big((b+1)/2\big)m_4+m_5-C_{a+b}, & b \text{ odd}. \end{cases}$$

And so on.

Note that

$$C_0 - \tfrac{1}{2}m_1 = C_0 - \tfrac{1}{2}am_2 - \tfrac{1}{2}m_3 = \begin{cases} C_a - \tfrac{1}{2}m_3, & a \text{ even};\\ \tfrac{1}{2}m_2 - (C_a - \tfrac{1}{2}m_3), & a \text{ odd}. \end{cases}$$

Hence we have the following rather curious rule for evaluating the number c of Theorem 1: Look at the odd numbered elements $a_1, a_3, \ldots, a_{t-1}$ of the canonical sequence for (h, k) and strike out all the even quotients in this list. If the remaining partial quotients are $a_{2j(0)+1}, a_{2j(1)+1}, \ldots, a_{2j(u-1)+1}$ for $0 \leqslant j(0) < \ldots < j(u-1) < t/2$, $u \geqslant 0$, the value of c is

$$\tfrac{1}{2}\big(m_1 + m_{2j(0)+2} - m_{2j(1)+2} + \ldots + (-1)^{u-1}m_{2j(u-1)+2}\big). \tag{3.23}$$

The following algorithm evaluates this formula.

Algorithm 2. Let h, k be relatively prime, $0 < h < k$. This algorithm will output the unique value of $c \leqslant \frac{1}{2}k$ which maximizes $\sigma(h, k, c)$ **for** $0 < c < k$.

```
procedure maxc (integer value $h, k$);
  begin integer $a, r, sigma$;
    $s := 1$; $sigma := h$;
    while $h > 0$ do
      begin comment At this point we have $k = m_{2j}$, $h = m_{2j+1}$
                    for some $0 \leqslant j < \lfloor t/2 \rfloor$, and $s$ is the sign of the
                    next term to be added in (3.23);
        $r := k \bmod h$; $k := h$; $h := r$;
        if $h = 0$
        then begin comment $t = 2j+1$, convert to $2j+2$;
            $sigma := sigma + s$;
          end
        else begin $a := \lfloor k/h \rfloor$; comment $a = a_{2j+1}$;
            if $a \bmod 2 = 1$ then
            begin $sigma := sigma + s \times h$; $s := -s$ end;
            $r := k \bmod h$; $k := h$; $h := r$;
         end;
      end;
    output ($sigma/2$);
  end.
```

The value of *sigma* remains positive and $\leqslant k$ throughout the algorithm. ■

4. Estimates for Dedekind sums. Our goal in this section is to obtain tight bounds on $|\sigma(h,k,c)|$ in terms of the canonical sequence of partial quotients $a_0, a_1, \ldots, a_{t-1}$ for (h,k), where t is even (cf. (3.18)). Throughout this section h, k are relatively prime, $0 < h < k$, and c is an integer.

We have proved in (2.14) that

$$\sigma(h,k,0) = \frac{h-p_{t-1}}{k} + \sum_{0\leqslant j<t} (-1)^j a_j. \tag{4.1}$$

It is well known from the theory of continued fractions (cf. [10], eq. 4.5.3–8) that

$$p_{t-1}h \equiv -1 \,(\text{modulo}\, k). \tag{4.2}$$

Hence if h' is the inverse of h modulo k, i.e.,

$$hh' \equiv 1 \,(\text{modulo}\, k) \quad \text{and} \quad 0 < h' < k, \tag{4.3}$$

we have $p_{t-1} = k-h'$, and (4.1) takes the more symmetrical form

$$\sigma(h,k,0) = \frac{h+h'}{k} - 1 + \sum_{0\leqslant j<t} (-1)^j a_j. \tag{4.4}$$

(Note that since $k = p_t = a_{t-1}p_{t-1}+p_{t-2}$, we have $h' \leqslant \frac{1}{2}k$ if and only if $a_{t-1} = 1$ in the canonical sequence.)

In order to estimate $\sigma(h,k,c)$ for $c \neq 0$ we shall first determine the sequences $b_0, b_1, \ldots, b_{t-1}$ and $c_0, c_1, \ldots, c_{t-1}$ defined in connection with Algorithm 1, for the special value c of Theorem 1 and equation (3.23). It is not difficult to prove that

$$b_{2j} = 0, \quad c_{2j} = c_{2j+1}; \tag{4.5}$$

$$b_{2j+1} = \tfrac{1}{2}a_{2j+1}, \quad \text{if} \quad a_{2j+1} \text{ is even;} \tag{4.6}$$

$$b_{2j(r)+1} = \tfrac{1}{2}\big(a_{2j(r)+1}+(-1)^r\big), \quad \text{for} \quad 0 \leqslant r < u; \tag{4.7}$$

$$c_{2j+1} = \tfrac{1}{2}\Big(m_{2j+1} + \sum_{\substack{0\leqslant r<u\\ j(r)\geqslant j}} (-1)^r m_{2j(r)+2}\Big); \tag{4.8}$$

for these values satisfy $c_j = b_j m_{j+1} + c_{j+1}$, $0 \leqslant c_{j+1} < m_{j+1}$, with one exception. The exception occurs when $a_{t-1} = 1$ and u is odd; for then $2j(u-1)+1 = t-1$, and $c_{t-2} = \frac{1}{2}(m_{t-1}+m_t) = 1$, and b_{t-2} should be 1, and c_{t-1} should be 0. We shall use equations (4.5)–(4.8) even in this exceptional case, and take care of the exception later.

The proof we shall discuss can be expressed very compactly in terms of $\sum$ notation, etc., but such a derivation would be quite hard to understand without some indication of how it could have been discovered. Indeed, equations (4.5)–(4.8) are not very easy to conceptualize until an example has been written down. Therefore, it will be helpful to consider an example in which $t = 10$; a_1, a_5, and a_9 are odd; a_3 and a_7 are even. We have the following tableau:

$$\begin{aligned}
m_0 &= a_0m_1+m_2, \qquad & m_1 &= a_1m_2+m_3,\\
&& c_0 &= c_1 = \tfrac{1}{2}(a_1+1)m_2+c_2 = \tfrac{1}{2}(m_1+m_2-m_6+m_{10});\\
m_2 &= a_2m_3+m_4, & m_3 &= a_3m_4+m_5,\\
&& c_2 &= c_3 = \tfrac{1}{2}a_3m_4+c_4 = \tfrac{1}{2}(m_3-m_6+m_{10});\\
m_4 &= a_4m_5+m_6, & m_5 &= a_5m_6+m_7,\\
&& c_4 &= c_5 = \tfrac{1}{2}(a_5-1)m_6+c_6 = \tfrac{1}{2}(m_5-m_6+m_{10});\\
m_6 &= a_6m_7+m_8, & m_7 &= a_7m_8+m_9,\\
&& c_6 &= c_7 = \tfrac{1}{2}a_7m_8+c_8 = \tfrac{1}{2}(m_7+m_{10});\\
m_8 &= a_8m_9+m_{10}, & m_9 &= a_9m_{10}+m_{11},\\
&& c_8 &= c_9 = \tfrac{1}{2}(a_9+1)m_{10}+c_{10} = \tfrac{1}{2}(m_9+m_{10});\\
&& m_{10} &= 1, \quad m_{11} = 0, \quad c_{10} = 0.
\end{aligned}$$

According to (2.14) the major unknown term in the evaluation of $\sigma(h,k,c)$ is

$$\sum (-1)^j b_j(c_j+c_{j+1})p_j,$$

and since $b_{2j} = 0$ we can express $-4\sum(-1)^j b_j(c_j+c_{j+1})p_j$ as follows:

$$\begin{aligned}
&(a_1+1)p_1(m_1+m_2+m_3-2m_6+2m_{10})+\\
&+\quad a_3p_3(m_3+m_5-2m_6+2m_{10})+\\
&+(a_5-1)p_5(m_5-m_6+m_7+2m_{10})+\\
&+\quad a_7p_7(m_7+m_9+2m_{10})+\\
&+(a_9+1)p_9(m_9+m_{10}).
\end{aligned}$$

We can rearrange these terms into $S+T$, where

$$S = a_1p_1(m_1+m_3)+a_3p_3(m_3+m_5)+a_5p_5(m_5+m_7)+a_7p_7(m_7+m_9)+a_9p_9m_9$$

is the portion which will be present whenever $t = 10$, regardless of the evenness or oddness of $a_1, a_3, \ldots, a_9$. Fortunately this sum turns out to be simply

$$S = (a_1+a_3+a_5+a_7+a_9)m_0 - m_1, \tag{4.9}$$

as we shall see in Lemma 5 below.

The remaining sum T can be written

$$\begin{aligned} T = {} & p_1(m_1+m_3+m_2-2m_6+2m_{10})+a_1p_1(m_2-2m_6+2m_{10})+\\ & \qquad +a_3p_3(-2m_6+2m_{10})-\\ & -p_5(m_5+m_7-m_6+2m_{10})+a_5p_5(-m_6+2m_{10})+\\ & \qquad +a_7p_7(2m_{10})+\\ & +p_9(m_9+m_{10})+a_9p_9(m_{10})\\ = {} & p_1(2m_3+m_2-2m_6+2m_{10})+a_1p_1(2m_2-2m_6+2m_{10})+\\ & \qquad +a_3p_3(-2m_6+2m_{10})-\\ & -p_5(2m_7-m_6+2m_{10})+a_5p_5(-2m_6+2m_{10})+\\ & \qquad +a_7p_7(2m_{10})+\\ & +p_9(2m_{11}+m_{10})+a_9p_9(2m_{10}). \end{aligned}$$

Now $a_jp_j = p_{j+1}-p_{j-1}$, so we get some helpful telescoping:

$$\begin{aligned} T = {} & 2m_2(p_2-p_0)-2m_6(p_6-p_0)+2m_{10}(p_{10}-p_0)+\\ & +2m_3p_1-2m_7p_5+2m_{11}p_9+\\ & +m_2p_1+m_6(p_5-2p_1)+m_{10}(p_9-2p_5+2p_1). \end{aligned}$$

Furthermore, $m_0 = m_2p_2+m_3p_1 = m_6p_6+m_7p_5 = m_{10}p_{10}+m_{11}p_9$ by (2.11), so

$$T = 2m_0+m_2(p_1-2)+m_6(p_5-2p_1+2)+m_{10}(p_9-2p_5+2p_1-2). \tag{4.10}$$

The coefficients of m_6 and m_{10} are nonnegative since $p_{j+2} \geqslant \frac{1}{2}p_j$ and $p_{j+1} \geqslant p_j$ for $0 \leqslant j < t$; hence

$$T \geqslant 2m_0+m_2(a_0-2). \tag{4.11}$$

Finally $m_jp_{j-1} = m_j(p_j-p_{j-2})/a_{j-1} \leqslant m_jp_j/a_{j-1} \leqslant m_0/a_{j-1}$, so

$$T \leqslant m_0(2+1/a_1+1/a_5+1/a_9). \tag{4.12}$$

(Note that if $a_0 = a_2 = \ldots = a_8 = x$, then

$$T/m_0 \to 2+1/a_1+1/a_5+1/a_9 \quad \text{as} \quad x\to\infty;$$

i.e., the upper bound in (4.12) is sharp.)

Let us now prove that the analog of (4.9) holds in general.

LEMMA 5. *If $m_j = a_jm_{j+1}+m_{j+2}$ and $p_{j+1} = a_jp_j+p_{j-1}$ for $0 \leqslant j < t$, and $p_0 = 1$, $p_{-1} = 0$, $m_t = 1$, $m_{t+1} = 0$, t even, then*

$$m_1+\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{odd}}} a_jp_j(m_j+m_{j+2}) = m_0\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{odd}}} a_j, \tag{4.13}$$

$$\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{even}}} a_jp_j(m_j+m_{j+2}) = p_{t-1}+m_0\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{even}}} a_j. \tag{4.14}$$

Proof. Since $p_jm_j+p_{j-1}m_{j+1} = m_0$, we have

$$\begin{aligned} a_jp_j(m_j+m_{j+2}) &= a_jm_0-a_jp_{j-1}m_{j+1}+a_jp_jm_{j+2}\\ &= a_jm_0-p_{j-1}(m_j-m_{j+2})+(p_{j+1}-p_{j-1})m_{j+2}\\ &= a_jm_0-p_{j-1}m_j+p_{j+1}m_{j+2} \end{aligned}$$

for $0 \leqslant j < t$. Hence the sums on the left of (4.13) and (4.14) are immediately evaluated. ∎

We are now ready to prove our main result.

THEOREM 2. *Let h and k be relatively prime, $0 < h < k$, and let c be an integer. Let $a_0, a_1, \ldots, a_{t-1}$ be the canonical sequence of partial quotients for (h, k), t even* (cf. (3.18)). *Then*

$$\sigma(h,k,c) \leqslant \Big(\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{even}}} a_j\Big)+\Big(\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{odd}}} \tfrac{1}{2}a_j\Big)-\tfrac{1}{2}. \tag{4.15}$$

Moreover, there exists a value of c for which

$$\sigma(h,k,c) \geqslant \Big(\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{even}}} a_j\Big)+\Big(\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{odd}}} \tfrac{1}{2}a_j\Big)-4-\tfrac{3}{2}\Big(\sum_{\substack{0\leqslant j<t\\ j\ \mathrm{odd\ and}\\ a_j\ \mathrm{odd}}} 1/a_j\Big). \tag{4.16}$$

Proof. By (2.6) we have

$$\sigma(h,k,c) = \sigma(h,k,0)-3(-1)^z+6\sum_{0\leqslant j<t}(-1)^j\Big(\frac{c_j^2}{m_jm_{j+1}}-b_j\Big)$$

for $0 < c < k$, where z is the least subscript such that $c_z = 0$. The value (3.23) of c which maximizes $\sigma(h,k,c)$ always has z even except when $a_{t-1} = 1$ and u is odd (cf. the discussion following (4.8)). It follows that

$$\max_{0<c<k}\sigma(h,k,c) = \sigma(h,k,0)-3+6\sum_{0\leqslant j<t}(-1)^j\Big(\frac{c_j^2}{m_jm_{j+1}}-b_j\Big) \tag{4.17}$$

holds without exception when the b_j and c_j are defined by (4.5)–(4.8). Now

$$\begin{aligned} -6\sum_{0\leqslant j<t}(-1)^jb_j &= 3\sum_{0\leqslant j<t/2}a_{2j+1}+3\sum_{0\leqslant r<u}(-1)^r\\ &= 3\sum_{0\leqslant j<t/2}a_{2j+1}+\tfrac{3}{2}\big(1-(-1)^u\big). \end{aligned} \tag{4.18}$$

The argument leading up to (4.10) proves in general that

$$(4.19)\qquad 6\sum_{0\leqslant j<t}(-1)^j c_j^2/m_j m_{j+1} = \frac{6}{m_0}\sum_{0\leqslant j<t}(-1)^j b_j(c_j+c_{j+1})p_j$$
$$= -\frac{3}{2}\sum_{0\leqslant j<t/2} a_{2j+1}+\frac{3}{2}\frac{m_1}{m_0}-\frac{3}{2}\big(1-(-1)^u\big)-\frac{3}{2}R,$$

$$(4.20)\qquad R = \frac{1}{m_0}\sum_{0\leqslant r<u} m_{2j(r)+2}\Big(p_{2j(r)+2}+2\sum_{0\leqslant s<r}(-1)^{r-s}p_{2j(s)+1}-2(-1)^r\Big).$$

As in (4.11) and (4.12) we conclude that

$$(4.21)\qquad \frac{h}{k}-1\leqslant\frac{m_2(a_0-2)}{m_0}\leqslant R\leqslant\sum_{0\leqslant r<u}1/a_{2j(r)+1}\leqslant u.$$

Note that $R = 0$ when $u = 0$; i.e., we have an exact result when $a_1, \ldots, a_{t-1}$ are all even. Combining (4.4) with (4.17), (4.18), and (4.19) now yields

$$(4.22)\qquad \max_{0<c<k}\sigma(h,k,c) = \Big(\sum_{\substack{0\leqslant j<t\\ j\text{ even}}} a_j\Big)+\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ odd}}}\tfrac{1}{2}a_j\Big)+$$
$$+\frac{3}{2}\frac{h}{k}+\frac{h+h'}{k}-\frac{3}{2}R-4.$$

Since (4.15) is easily verified when $c = 0$, the proof of (4.15) and (4.16) is immediate. ∎

It is amusing to note that when $a_{2j} = x$ and $a_{2j+1} = 2y$ for all j, we have $h' = k-xh/2y$, and there is a simple explicit formula

$$\max_{0\leqslant c<k}\ \sigma(h,k,c) = \tfrac{1}{2}t(x+y)+\tfrac{1}{2}(5-x/y)h/k-3.$$

Theorem 2 can also be used to obtain bounds on the minimum value of $\sigma(h,k,c)$:

THEOREM 3. *Under the assumptions of Theorem 2,*

$$(4.23)\qquad \sigma(h,k,c)\geqslant-\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ even}}}\tfrac{1}{2}a_j\Big)-\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ odd}}}a_j\Big)+\tfrac{1}{2}.$$

Moreover, there exists a value of c for which

$$(4.24)\qquad \sigma(h,k,c)\leqslant-\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ even}}}\tfrac{1}{2}a_j\Big)-\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ odd}}}a_j\Big)+4+\tfrac{3}{2}\Big(\sum_{\substack{0\leqslant j<t\\ j\text{ even and}\\ a_j\text{ odd}}}1/a_j\Big).$$

Proof. Since $k-h' = p_{t-1}$, the canonical sequence for $(k-h', k)$ is

$$(4.25)\qquad a_{t-1}, a_{t-2}, \ldots, a_0.$$

Note that even and odd positions are interchanged here (as are the p's and the m's). Now

$$(4.26)\qquad \sigma(k-h',k,h'c) = -12\sum_{0\leqslant j<k}\Big(\Big(\frac{j}{k}\Big)\Big)\Big(\Big(\frac{h'j-h'c}{k}\Big)\Big)$$
$$= -12\sum_{0\leqslant j<k}\Big(\Big(\frac{j+c}{k}\Big)\Big)\Big(\Big(\frac{h'j}{k}\Big)\Big)$$
$$= -12\sum_{0\leqslant j<k}\Big(\Big(\frac{hj+c}{k}\Big)\Big)\Big(\Big(\frac{h'hj}{k}\Big)\Big) = -\sigma(h,k,c),$$

hence maximizing $\sigma(k-h',k,c)$ is equivalent to minimizing $\sigma(h,k,c)$ and changing the sign. ∎

COROLLARY. *Under the hypotheses of Theorem 2,*

$$(4.27)\qquad |\sigma(h,k,c)|\leqslant\sum_{0\leqslant j<t}a_j-\tfrac{1}{4}(t+2).$$

Proof. In fact, by combining (4.15) and (4.23) we have

$$|\sigma(h,k,c)|\leqslant\Big(\sum_{0\leqslant j<t}a_j\Big)-\tfrac{1}{2}-\tfrac{1}{2}\min(a_0+a_2+\ldots+a_{t-2},$$
$$a_1+a_3+\ldots+a_{t-1}).\ ∎$$

A combination of (4.16) and (4.23) yields

$$(4.28)\qquad \max_c\sigma(h,k,c)-\min_c\sigma(h,k,c)\geqslant\tfrac{3}{2}\sum_{0\leqslant j<t}\big(a_j-(a_j\bmod 2)/a_j\big)-8.$$

This bound is weakest (indeed, trivial) when we have the Fibonacci case $a_0 = a_1 = \ldots = a_{t-1} = 1$, $h = F_t$, $k = F_{t+1}$ where

$$(4.29)\qquad F_0 = 0,\quad F_1 = 1,\quad F_{j+2} = F_{j+1}+F_j;$$

so it will be of interest to examine $\max(F_t, F_{t+1}, c)-\min(F_t, F_{t+1}, c)$. Both max and min have the same magnitude whenever $(a_0, a_1, \ldots, a_{t-1}) = (a_{t-1}, a_{t-2}, \ldots, a_0)$, i.e., whenever $h' = k-h$ (cf. the proof of Theorem 3), hence it suffices to consider $\max(F_t, F_{t+1}, c)$. By (4.22) we have

$$(4.30)\qquad \max_{0<c<F_{t+1}}\ \sigma(F_t, F_{t+1}, c) = \tfrac{3}{4}t+\tfrac{3}{2}\frac{h}{k}-\tfrac{3}{2}R-3,$$

where

$$(4.31)\qquad R = \frac{1}{F_{t+1}}\sum_{0\leqslant r<t/2}F_{t-2r-1}\Big(F_{2r+2}+2\sum_{0\leqslant s<r}(-1)^{r-s}F_{2s+2}-2(-1)^r\Big).$$

For example, when $t = 8$,

$$R = \frac{1}{F_9}\big(F_7(F_2-2)+F_5(F_4-2F_2+2)+F_3(F_6-2F_4+2F_2-2)+ \\ +F_1(F_8-2F_6+2F_4-2F_2+2)\big).$$

Let $L_n = F_{n+1}+F_{n-1}$. Using the easily verified identities

$$(4.32)\qquad F_{2r}-2F_{2r-2}+2F_{2r-4}+\ldots+(-1)^{r-1}2F_2+(-1)^r2 \\ = \tfrac{1}{5}\big(L_{2r}+(-1)^r8\big),$$

$$(4.33)\qquad F_{2n-1}L_2+F_{2n-3}L_4+\ldots+F_1L_{2n} = nF_{2n+1}+F_{2n},$$

we find

$$\begin{aligned} R &= \frac{1}{5F_9}\big(F_7(L_2-8)+F_5(L_4+8)+F_3(L_6-8)+F_1(L_8+8)\big) \\ &= \frac{1}{5F_9}\big(4F_9+F_8-8(F_8-2F_6+2F_4-2F_2)\big) \\ &= \frac{1}{5F_9}\left(4F_9+F_8-8\left(\frac{1}{5}\,(L_8+8)-2\right)\right), \end{aligned}$$

and in general we obtain the exact formula

$$(4.34)\qquad \begin{aligned} R &= \frac{1}{5F_{t+1}}\left(\frac{t}{2}F_{t+1}+F_t-8\left(\frac{1}{5}\,(L_t+(-1)^{t/2}8)-(-1)^{t/2}2\right)\right) \\ &= \frac{1}{10}t-\frac{16}{25}+\frac{13}{25}\,\frac{h}{k}+\frac{16}{25}\,\frac{(-1)^{t/2}}{k} \end{aligned}$$

when $h = F_t$ and $k = F_{t+1}$, t even.

Instead of using a strict Euclidean algorithm to compute $\sigma(h, k, c)$, it is possible to use the so-called *least remainder algorithm*, which replaces h by $k-h$, if necessary, to ensure that $h \leqslant \frac{1}{2}k$ at each step. The least remainder algorithm is a combination of the additive and subtractive processes, and it can be obtained from the additive process as follows: When $a_j = 1$ and $a_{j-1} > 1$, replace

$$(4.35)\qquad \begin{array}{ll} & \text{by} \\ m_{j-1} = a_{j-1}m_j+m_{j+1} & m_{j-1} = (a_{j-1}+1)m_j-m_{j+2} \\ m_j = m_{j+1}+m_{j+2}. & m_j = (a_{j+1}+1)m_{j+2}+m_{j+3} \\ m_{j+1} = a_{j+1}m_{j+2}+m_{j+3}. & \end{array}$$

This saves one iteration for each partial quotient 1 that is immediately preceded by an even number of partial quotients equal to 1, and it is known ([10], exercise 4.5.3–29) that the number of iterations decreases by about $2-\log_2(1+\sqrt{5}) \approx 30\%$ on the average. The transformation

increases $\sum a_j$ by 1, so it increases the bound (4.27). Other changes from an additive to a subtractive procedure also increase the bound, except when $a_{t-1} > 1$ is replaced by $(a_{t-1}-1, 1)$. In other words:

Corollary. *The bound*

$$(4.36)\qquad |\sigma(h, k, c)| \leqslant \sum_{0\leqslant j<t} a_j - \tfrac{1}{4}t$$

holds for all sequences of positive integers $a_0, a_1, \ldots, a_{t-1}$ such that $m_0 = k$, $m_1 = h$, $m_j = a_jm_{j+1}\pm m_{j+2}$, $m_t = 1$, and $m_{t+1} = 0$; in particular, it holds for the least remainder algorithm.

Proof. Let θ be any number between 0 and $1/3$; we will prove that the minimum value of $\sum a_j - \theta t$, over all sequences $a_0, \ldots, a_{t-1}$ as described in the corollary, occurs when the a_j are defined by the additive Euclidean algorithm modified so that $a_{t-1} = 1$.

It is not quite easy to prove this statement rigorously, as the reader will see if he makes an attempt, since there is no obvious quantity which can be used as the basis of a valid proof by induction. In fact, the result would be false for $1/3 < \theta < 1/2$, although this is not immediately evident. For example let $h = 2$, $k = 7$; the modified Euclidean algorithm has $(a_0, \ldots, a_{t-1}) = (3, 1, 1)$ and the sum is $5-3\theta$, but the sequence $(a_0, \ldots, a_{t-1}) = (1, 1, 1, 1, 1, 1)$ has a sum of $6-6\theta$.

Our approach will be to consider an infinite directed graph on the vertices (h, k), for all pairs h, k of nonnegative, relatively prime integers. The arcs of this directed graph will go from (h, k) to $(|k-ah|, h)$ for all positive integers a, and every such arc will be assigned a "distance" $a-\theta$. In order to prove that the modified Euclidean algorithm gives the "shortest path" from (h, k) to $(0, 1)$ for all h and k, it suffices to prove that $f(h, k) \leqslant a-\theta+f(|k-ah|, h)$ for all $a \geqslant 1$, when $f(h, k)$ is the distance to $(0, 1)$ in the modified Euclidean algorithm.

Let $f(0, 1) = 0$, $f(1, 1) = 1-\theta$, $f(1, k) = k-2\theta$ for $k \geqslant 2$, $f(h, h) = 1-\theta+f(h-k, h)$ for $h > k$, and $f(h, k) = \lfloor k/h\rfloor-\theta+f(k \bmod h, h)$ for $1 < h < k$. It follows that

$$f(h, k) = f(k, h)+\begin{cases} 1, & \text{if}\quad h > k > \frac{1}{2}h, \\ 1-\theta, & \text{if}\quad \frac{1}{2}h = k, \\ 1-2\theta, & \text{if}\quad \frac{1}{2}h > k \end{cases}$$

and

$$f(h, ah-k) = a-2+f(h, k),\qquad \text{if}\quad h > k \text{ and } a \geqslant 2.$$

We must prove that $f(h, k) \leqslant a-\theta+f(|k-ah|, h)$ for all $a \geqslant 1$. This inequality is readily verified for $h = 1$. When $1 < h < k$, let $b = \lfloor k/h\rfloor$, so that equality holds for $a = b$. If $1 \leqslant a < b$, we have $f(|k-ah|, h)$

$\geqslant 1-2\theta+f(h, k-ah) = 1-2\theta+f(h, k)-a \geqslant f(h, k)-a+\theta$ since $\theta \leqslant 1/3$; if $a = b+1$, we have $f(|k-ah|, h) = f\big(h-(k \bmod h), h\big) = f(h, k \bmod h) - 1+\theta \geqslant 1-2\theta+f(k \bmod h, k)-1+\theta = f(h, k)-a+1 > f(h, k)-a+\theta$; and if $a > b+1$, we have

$$\begin{aligned} f(|k-ah|, h) &= f\big((a-b)h-(k \bmod h), h\big) \geqslant 1-2\theta+f\big(h, (a-b)h-(k \bmod h)\big) \\ &= 1-2\theta+(a-b-2)+f(h, k \bmod h) \\ &\geqslant 1-2\theta+(a-b-2)+1-2\theta+f(k \bmod h, h) \\ &= a-2b-3\theta+f(h, k) > -a+\theta+f(h, k). \end{aligned}$$

Finally if $h > k$ and $a \geqslant 2$ we have

$$\begin{aligned} f(|k-ah|, h) &= f(ah-k, h) \geqslant 1-2\theta+f(h, ah-k) \\ &= a-1-2\theta+f(h, k) > f(h, k)-a+\theta. \ ■ \end{aligned}$$

The inequality (4.36) is a slight improvement on the results of U. Dieter and J. Ahrens ([7], Theorem 4.8) who showed that

$$|\sigma(h, k, c)| \leqslant \sum a_j + 3t + 5.$$

For applications to random number generation, we would like to know that $\sigma(h, k, c)$ is not too large. A. Khintchine has shown [8] that for all $\varepsilon > 0$ the measure of the set of real numbers with

$$\left|\frac{a_0+\ldots+a_{n-1}}{n\log_2 n} - 1\right| > \varepsilon \tag{4.37}$$

approaches zero as $n \to \infty$. Hence by applying Lemma 4.5.3M of [10] we can show that (for sufficiently large fixed n and all large k) the number of values of h whose first n partial quotients satisfy (4.37) is less than εk. This is not as strong a result as one would like, but it does suggest that the *average* sum of partial quotients satisfies

$$\frac{1}{k}\sum_{1\leqslant h\leqslant k}\Big(\sum a_j\Big) \leqslant C(\log k)(\log\log k) \tag{4.38}$$

for some appropriate constant C. (It is well known that $t \leqslant \log_\phi k$, where $\phi = (1+\sqrt{5})/2$, cf. [10], Theorem 4.5.3L.)

Yao and Knuth [14] have recently established the somewhat surprising fact that (4.38) is false; in fact,

$$\frac{1}{k}\sum_{1\leqslant h\leqslant k}\Big(\sum a_j\Big) = \frac{6(\ln k)^2}{\pi^2} + O\big((\log k)(\log\log k)^2\big) \tag{4.39}$$

as $k \to \infty$. Apparently the "middle" partial quotients tend to be larger than the first ones.

The results in [14] imply that the average of $\sum a_j$ restricted to odd values of j is asymptotically $3(\ln k)^2/\pi^2$; thus the value of our bound on $|\sigma(h, k, c)|$ is $\sim\frac{9}{2}(\ln k)^2/\pi^2$ for fixed k, when averaged over $1 \leqslant h \leqslant k$. It follows that at most $O\big((\log k)^{-\varepsilon}\big)$ choices of h will have $|\sigma(h, k, c)| > (\log k)^{2+\varepsilon}$. This supports the empirically observed phenomenon that "random" choices of h almost always lead to satisfactory random number generators.

5. A general reciprocity law.

It remains for us to prove Lemma 3 in the case $c = 0$. Let us consider first an extremely general identity:

LEMMA 6. *Let $f(x), g(x)$ be any real-valued functions defined over the nonnegative integers, and let m, n be positive integers. Let a be any positive real number. Then*

$$\begin{aligned} \sum_{0\leqslant j<an}\big(f(j+1)-f(j)\big)g(\lfloor mj/n\rfloor+1) + \sum_{0\leqslant r<am} f(\lceil rn/m\rceil)\big(g(r+1)-g(r)\big) \\ = f(\lceil an\rceil)g(\lceil am\rceil)-f(0)g(0). \end{aligned} \tag{5.1}$$

Proof. Consider the change of variable $r = \lfloor mj/n\rfloor$, a condition which holds if and only if

$$\begin{gathered} r \leqslant mj/n < r+1, \\ \frac{rn}{m} \leqslant j < \frac{(r+1)n}{m}, \\ \left\lceil\frac{rn}{m}\right\rceil \leqslant j < \left\lceil\frac{(r+1)n}{m}\right\rceil. \end{gathered}$$

This range of values of j is used for those r with

$$\frac{(r+1)n}{m} < an;$$

the next value of r satisfies

$$\frac{rn}{m} \leqslant j < an \leqslant \frac{(r+1)n}{m}, \quad \text{i.e.,} \quad r = \lceil am\rceil - 1.$$

Hence

$$\begin{aligned} &\sum_{0\leqslant j<an}\big(f(j+1)-f(j)\big)g(\lfloor mj/n\rfloor+1) \\ &\qquad = \sum_{0\leqslant r<am-1} g(r+1)\big(f(\lfloor(r+1)n/m\rceil)-f(\lceil rn/m\rceil)\big) + \\ &\qquad\qquad + g(\lceil am\rceil)\big(f(\lceil an\rceil)-f(\lceil(\lceil am\rceil-1)n/m\rceil)\big). \end{aligned}$$

Rearranging the latter sum by grouping terms with the same value of $f(\lceil rn/m\rceil)$ yields the result. ■

(Stieltjes integration by parts can be used to give another proof of (5.1) and formulas of even greater generality.)

COROLLARY.

$$(5.2)\qquad \sum_{0\leqslant j<an}\binom{j}{q}\binom{\lfloor mj/n\rfloor+1}{p+1}+\sum_{0\leqslant j<am}\binom{\lceil jn/m\rceil}{q+1}\binom{q}{p}=\binom{\lceil an\rceil}{q+1}\binom{\lceil am\rceil}{p+1}.$$

Proof. Set $f(x)=\binom{x}{q+1}$, $g(x)=\binom{x}{p+1}$ in (5.1). ■

The general reciprocity law of Lemma 6, with $a=1/2$, $q=0$, $p=0$, lies at the heart of Eisenstein's proof of the law of quadratic reciprocity for prime numbers (cf. [9], exercise 1.2.4–47, and [2]). We shall now show that it immediately yields the reciprocity law for Dedekind sums $\sigma(h,k,0)$.

Let $a=1$, $p=1$, $q=0$, $m=h$, $n=k$ in (5.2), and express $\lfloor\ \rfloor$ and $\lceil\ \rceil$ in terms of $((\ ))$. Assuming that h and k are relatively prime, we have

$$\begin{aligned}kh(h-1)&=\sum_{0<j<k}\left(\frac{hj}{k}-\left(\left(\frac{hj}{k}\right)\right)+\frac{1}{2}\right)\left(\frac{hj}{k}-\left(\left(\frac{hj}{k}\right)\right)-\frac{1}{2}\right)+\\&\qquad+2\sum_{0<j<h}\left(\frac{kj}{h}-\left(\left(\frac{kj}{h}\right)\right)+\frac{1}{2}\right)j\\&=\sum_{0<j<k}\left(\left(\frac{hj}{k}\right)^2+\left(\left(\frac{hj}{k}\right)\right)^2-\left(\left(\frac{hj}{k}\right)\right)\left(2h\left(\left(\frac{j}{k}\right)\right)+h\right)-\frac{1}{4}\right)+\\&\qquad+\sum_{0<j<h}\left(\frac{2kj^2}{h}-\left(\left(\frac{kj}{h}\right)\right)\left(2h\left(\left(\frac{j}{h}\right)\right)+h\right)+j\right)\\&=\sum_{0<j<k}\left(\left(\frac{hj}{k}\right)^2+\left(\frac{j}{k}-\frac{1}{2}\right)^2-2h\left(\left(\frac{hj}{k}\right)\right)\left(\left(\frac{j}{k}\right)\right)-\frac{1}{4}\right)+\\&\qquad+\sum_{0<j<h}\left(\frac{2kj^2}{h}-2h\left(\left(\frac{kj}{h}\right)\right)\left(\left(\frac{j}{h}\right)\right)+j\right).\end{aligned}$$

Everything can now be summed, and we obtain the desired law,

$$kh^2-kh+\frac{1}{6}\,\frac{h^2}{k}+\frac{1}{6}\,k+\frac{1}{6k}-\frac{1}{2}\,h-\frac{1}{6}\,h\big(\sigma(h,k,0)+\sigma(k,h,0)\big)=kh(h-1).$$

It is important to note that the existence of a reciprocity formula connecting $f(h,k)$ with $g(k,h)$ does not necessarily imply that we have an efficient "Euclidean" algorithm for the evaluation of f and g; it is also necessary to have relations between $f(h\bmod k,k)$, $g(k\bmod h,h)$ and $f(h,k)$, $g(k,h)$.

For example, let us try to extend the above derivation to develop a reciprocity formula for cubic analogs of Dedekind sums, by taking $p=2$ in (5.2). A derivation like that above yields

$$(5.3)\qquad \sum_{0<j<k}\left(h^2\left(\left(\frac{j}{k}\right)\right)^2\left(\left(\frac{hj}{k}\right)\right)-h\left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{hj}{k}\right)\right)^2\right)+ +\sum_{0<j<h}h^2\left(\left(\frac{j}{h}\right)\right)^2\left(\left(\frac{kj}{h}\right)\right)=\varphi(h,k),$$

where $\varphi(h,k)$ can be explicitly evaluated in terms of h, k and Dedekind sums. However, this is a rather pointless identity, because the substitution $j\to k-j,\ h-j$ shows that both sides of (5.3) are zero! Turning to $p=3$ we find

$$(5.4)\qquad \sum_{0<j<h}\left(\frac{1}{6}\,k^3\left(\left(\frac{j}{h}\right)\right)^3\left(\left(\frac{kj}{h}\right)\right)-\frac{1}{4}\,k^2\left(\left(\frac{j}{h}\right)\right)^2\left(\left(\frac{kj}{h}\right)\right)^2+\frac{1}{6}\,k\left(\left(\frac{j}{h}\right)\right)\left(\left(\frac{kj}{h}\right)\right)^3\right)+ +\sum_{0<j<k}\frac{1}{6}\,k^3\left(\left(\frac{j}{k}\right)\right)^3\left(\left(\frac{hj}{k}\right)\right)=\psi_1(h,k),$$

but this equation by itself does not imply an efficient evaluation procedure. If we set $q=1$, $p=2$ we get an independent reciprocity formula,

$$(5.5)\quad \sum_{0<j<h}\left(\frac{1}{2}\,hk^2\left(\left(\frac{j}{h}\right)\right)^3\left(\left(\frac{kj}{h}\right)\right)-\frac{1}{2}\,hk\left(\left(\frac{j}{h}\right)\right)^2\left(\left(\frac{kj}{h}\right)\right)^2+\frac{1}{6}\,h\left(\left(\frac{j}{h}\right)\right)\left(\left(\frac{kj}{h}\right)\right)^3\right)+ +\sum_{0<j<k}\left(\frac{1}{2}\,hk^2\left(\left(\frac{j}{k}\right)\right)^3\left(\left(\frac{hj}{k}\right)\right)-\frac{1}{4}k^2\left(\left(\frac{j}{k}\right)\right)^2\left(\left(\frac{kj}{k}\right)\right)^2\right)=\psi_2(h,k).$$

Let

$$(5.6)\qquad \sigma_{mn}(h,k)=\sum_{0<j<k}\left(\left(\frac{j}{k}\right)\right)^m\left(\left(\frac{hj}{k}\right)\right)^n.$$

Equations (5.4) and (5.5) combine to give an efficient procedure:

THEOREM 4. *There is an algorithm which computes $\sigma_{13}(h,k)$, $\sigma_{22}(h,k)$, and $\sigma_{31}(h,k)$ in $O(\log k)$ arithmetic operations.*

Proof. By definition, $\sigma_{mn}(h,k)=\sigma_{mn}(h\bmod k,k)$. Equations (5.4) and (5.5) tell us that

$$\begin{aligned}\tfrac{1}{6}k^3\sigma_{31}(h,k)&=\psi_1(h,k)-\tfrac{1}{6}k^3\sigma_{31}(k,h)+\tfrac{1}{4}k^2\sigma_{22}(k,h)-\tfrac{1}{6}k\sigma_{13}(k,h),\\ \tfrac{1}{4}k^2\sigma_{22}(h,k)&=-\psi_2(h,k)+\tfrac{1}{2}hk^2\sigma_{31}(h,k)+\tfrac{1}{2}hk^2\sigma_{31}(k,h)-\\&\qquad-\tfrac{1}{2}hk\sigma_{22}(k,h)+\tfrac{1}{6}h\sigma_{13}(k,h),\\ \tfrac{1}{6}k\sigma_{13}(h,k)&=\psi_2(k,h)+\tfrac{1}{2}hk\sigma_{22}(h,k)-\tfrac{1}{2}h^2k\sigma_{31}(h,k)-\\&\qquad-\tfrac{1}{2}h^2k\sigma_{31}(k,h)+\tfrac{1}{4}h^2\sigma_{22}(k,h).\end{aligned}$$

Therefore a Euclidean algorithm applies. ■

A similar argument shows that we can evaluate any $\sigma_{mn}(h, k)$ in $O\big((m+n)^3 \log k\big)$ operations.

Reciprocity laws for sums of polynomials such as (5.6), but in a completely different notation, have previously been obtained by T. M. Apostol [1] and L. Carlitz [3].

In order to carry the application to random-number generators further, it will be neces ary to deal with sums of a still more general type, e.g.

$$\sum_{0<j<k} \left(\left(\frac{j}{k}\right)\right)\left(\left(\frac{h_1 j + c_1}{k}\right)\right)\left(\left(\frac{h_2 j + c_2}{k}\right)\right).$$

Reciprocity laws for such sums (even if we had them only in the special cases $h_2 = 1$, $h_2 = h_1$, $h_2 = h_1^2$) would be useful for further development of the theory.

Acknowledgment. The MACSYMA system was very useful not only in developing Algorithm 1, but also for much of the formula manipulation and experimental computations as the theorems were developed. MACSYMA's exact rational arithmetic and its ability to compute functions symbolically were of considerable utility in this research.

References

[1] T. M. Apostol, *Theorems on generalized Dedekind sums*, Pacific J. Math 2 (1952), pp. 1–9.

[2] B. C. Berndt, *A generalization of a theorem of Gauss on sums involving* $[x]$, Amer. Math. Monthly 82 (1975), pp. 44–51.

[3] L. Carlitz, *A reciprocity and four-term relation for generalized Dedekind sums*, Indag. Math. 36 (1974), pp. 413–422.

[4] J. W. S. Cassels, *An Introduction to the Geometry of Numbers*, Springer, Berlin 1959, p. 34.

[5] Ulrich Dieter, *Das Verhalten der Kleinschen Funktionen* $\log \sigma_{g,h}(w_1, w_2)$ *gegenüber Modultransformationen und verallgemeinerte Dedekindsche Summen*, Journ. Reine Angew. Math. 201 (1959), pp. 37–70.

[6] U. Dieter and J. Ahrens, *An exact determination of serial correlations of pseudo-random numbers*, Numer. Math. 17 (1971), pp. 101–123.

[7] — — *Uniform Random Numbers*, Institut für Math. Statistik, Tech. Hoch. Graz, Austria, 1974; to be published by John Wiley and Sons.

[8] A. Khintchine, *Metrische Kettenbruchprobleme*, Compositio Math. 1 (1935), pp. 361–382.

[9] Donald E. Knuth, *The Art of Computer Programming, Fundamental Algorithms*, vol. 1, 2nd edition, Addison-Wesley, Reading, Mass., 1973.

[10] — *The Art of Computer Programming, Seminumerical Algorithms*, vol. 2, Addison-Wesley, Reading, Mass., 1969.

[11] The Mathlab Group, *MACSYMA Reference Manual*, Project MAC, Mass. Inst. of Technology, version six (1974).

[12] Hans Rademacher and Emil Grosswald, *Dedekind Sums*, **Math.** Assoc. Amer., Carus Monograph No. 16 (1972).

[13] Bernhard Riemann's gesammelte *Mathematische Werke*, ed. by Heinrich Weber, 2nd edition, 1892. (Reprinted by Dover Publications, New York, 1953.) See R. Dedekind, *Erläuterungen zu den Fragmenten* XXVIII.

[14] Andrew C. Yao and Donald E. Knuth, *Analysis of the subtractive algorithm for greatest common divisors*, Proc. Nat. Acad. Sci. 72 (1975), pp. 4720–4722.

Received on 27. 2. 1975
and in revised form on 14. 5. 1976 (682)