Proof of Theorem 2. Let $f(x) = a_0x^n + \ldots + a_n$ be a polynomial satisfying the assumptions of the theorem. Let α be any of its roots. By the assumptions, α is different from zero and is not a root of unity. Put $k_1 = Q(\alpha)$. By Theorem 1 we have (2), since $c_{k_1}(\alpha) = c(\alpha) = c(f)$. From the Theorem of [2] and the remark at the end of that paper it follows that the group G_2 is uniquely determined by the polynomial f and the positive integer k. Put

$C = \{\mathfrak{q}$: $\mathfrak{q}$ a prime ideal of k_1, $N\mathfrak{q} \equiv 1 \bmod k$, $N\mathfrak{q} \equiv r \bmod D$, α is a kth power residue $\bmod \mathfrak{q}\}$,

$B = \{q$: q a prime number, $q \equiv 1 \bmod k$, $q \equiv r \bmod D$, the congruence $f(x^k) \equiv 0 \bmod q$ is solvable$\}$,

where $(r, D) = 1$ and residue class of $r \bmod D$ contains a rational integer belonging to G_2.

By the same argument as in [1] we have

$$\frac{1}{n}d(C) \leqslant d(B) \leqslant \frac{1}{\varkappa}d(C). \tag{11}$$

By the definition of $c(f)$ and $C(f)$, $c(\alpha) = c(f)$, $C(\alpha) = C(f)$. Hence by Theorem 1

$$d(C) = \frac{n(k, c(f))}{C(f)k\varphi([D, k])}\frac{|K_1 \cap P_{[D,k]}|}{|K_1|}.$$

By (11)

$$\frac{(k, c(f))}{C(f)k\varphi([D, k])}\frac{|K_1 \cap P_{[D,k]}|}{|K_1|} \leqslant d(B) \leqslant \frac{n}{\varkappa}\frac{(k, c(f))}{C(f)k\varphi([D, k])}\frac{|K_1 \cap P_{[D,k]}|}{|K_1|}.$$

Theorem 2 is proved.

References

[1] J. Wójcik, *Contributions to the theory of Kummar extensions*, Acta Arith. 40 (1982), 155–174.

[2] — *On the congruence $f(x^k) \equiv 0 \bmod q$, where q is a prime and f is a k-normal polynomial*, ibid. 41 (1982), 151–161.

Received on 22.11.1985
and in revised form on 9.1.1987 (1563)

ACTA ARITHMETICA
LI (1988)

Multiplicative functions and Brun's sieve

by

KRISHNASWAMI ALLADI (Gainesville, Florida)

1. Introduction. Let g be a strongly multiplicative function. That is

$$g(n) = \prod_{\substack{p|n \\ p=\text{prime}}} g(p).$$

The *truncation* of g at y is

$$g_y(n) = \prod_{\substack{p|n \\ p<y}} g(p).$$

As is customary null products have value one.

For any set $\mathscr{A}$ of positive integers we let $\mathscr{A}(x)$ denote $\mathscr{A} \cap [1, x]$. The problem we consider here is the estimation of

$$S_g(\mathscr{A}(x), y) = \sum_{n\in\mathscr{A}(x)} g_y(n) \tag{1.1}$$

for sets $\mathscr{A}$ satisfying certain conditions to be specified in Section 2. We were motivated to study this sum because it turns out (as will be seen in Section 3) to be a natural generalization of a typical sieve problem. We show that Brun's sieve could be used to estimate $S_g(\mathscr{A}(x), y)$ when $-1 \leqslant g \leqslant 1$, provided $\alpha = (\log|\mathscr{A}(x)|)/\log y$ is not small (see § 5–§ 7) and for this we make use of an interesting 'monotonicity principle' (see § 4).

Previously [1], [2], [3] we had investigated such sums when $0 < g \leqslant 1$. In this case g may be written as

$$g(n) = e^{uf(n)}$$

where $u < 0$ and $f \geqslant 0$ is a strongly additive function. So the sum in (1.1) can be interpreted in terms of the Laplace transform of f_y, which is the truncation of f and y. Such an approach led to a new method of estimating the moments of f using the sieve. For the sake of completeness we shall state (without proof) towards the end of Section 6 some results for the case $0 \leqslant g \leqslant 1$ but in a slightly stronger form than was utilized by us earlier. The main interest in the present paper lies in showing that the sieve can be employed to deal with such sums even when g admits both positive and negative values and as far as we know this has not been done earlier.

When $\mathscr{A} = N$, the set of all positive integers, we use the notation $S_g(x, y)$ for the sum in (1.1). In Section 8 we just point out how two other methods could be used to estimate $S_g(x, y)$ even when g takes complex values because of the special properties of N. The first one which uses contour integration is effective when α is not small and $|g(p)|$ not large. The second which relies on a recurrence satisfied by $S_g(x, y)$ works well if $g(p)$ has an average value and α is not too large. In Section 9 a lower bound for $S_g(x) = S_g(x, x)$ that is uniform for $0 \leqslant g \leqslant 1$ is established and this is an extension of a small sieve inequality due to Erdős and Ruzsa [5].

All notation introduced so far will be retained. For instance g will always represent a strongly multiplicative function. The '$\ll$' and 'O' symbols are equivalent. Unless indicated otherwise implicit constants are either absolute or depend at most on the set $\mathscr{A}$ and this will be clear from the context. By $c_0, c_1, \ldots$ we mean positive constants whose values will not concern us. Finally, the letters p and q, with or without subscripts, will always denote primes.

2. Dual functions and special sets. Given g, consider the strongly multiplicative function g^* generated by $g^*(p) = 1-g(p)$. We may think of g, g^* as a pair of dual functions because

$$g^*(n) = \sum_{d|n} \mu(d)g(d) \quad \text{and} \quad g(n) = \sum_{d|n} \mu(d)g^*(d),$$

where μ is the Möbius function.

The dual g^* arises naturally while estimating $S_g(\mathscr{A}(x)) = S_g(\mathscr{A}(x), x)$. That is

$$S_g(\mathscr{A}(x)) = \sum_{n\in\mathscr{A}(x)} \sum_{d|n} \mu(d)g^*(d) = \sum_{d\leqslant x} \mu(d)g^*(d)|\mathscr{A}_d(x)|,$$

where $\mathscr{A}_d = \{n\in\mathscr{A} \mid n \equiv 0 \pmod d\}$. So, as in sieve theory, we are led to consider sets $\mathscr{A}$ for which the quantity $|\mathscr{A}_d(x)|$ satisfies certain conditions.

Here we shall assume that

$$|\mathscr{A}_d(x)| = \frac{X\omega(d)}{d} + R_d(x), \tag{2.1}$$

where $X = |\mathscr{A}(x)|$, ω is multiplicative and $0 \leqslant \omega(p) \ll 1$ for all p. Regarding $R_d(x)$ we assume that there exists $c_0 > 1$ such that

$$|R_d(x)| \ll \left(\frac{X\log X}{d} + 1\right) c_0^{\nu(d)}, \tag{2.2}$$

where $\nu(d) = \sum_{p|d} 1$. Moreover the $R_d(x)$ will be small on average in the sense that there is $\beta \in (0, 1]$ with the following property: For every $U > 0$, there is $V > 0$ such that

$$\sum_{d\leqslant X^\beta/\log^V X} |R_d(x)| \ll_U \frac{X}{\log^U X}. \tag{2.3}$$

Sets $\mathscr{A}$ satisfying the above conditions will be called *special*.

An example of a special set is

(E−1) $\quad \mathscr{A} = \{p+a \mid p = \text{prime}\}$, where a is a fixed integer.

Here we may take $\beta = 1/2$ and $c_0 = 1$. Many special sets satisfy the stronger condition

$$|R_d(x)| \ll \omega(d) \tag{2.4}$$

in which case

$$c_0 = \sup_p \omega(p), \quad \beta = 1 \quad \text{and} \quad V = U-1+c_0.$$

Condition (2.4) occurs for instance in the case of

(E−2) $\quad \mathscr{A} = \{F(n) \mid n = 1, 2, 3, \ldots\}$, where $F(x) \in N[x]$.

3. Truncated functions and the sieve. Let P denote a set of prime numbers and $P_y = \prod_{p\in P, p<y} p$. A typical sieve problem is to estimate

$$S^{(P)}(\mathscr{A}(x), y) = \sum_{\substack{n\in\mathscr{A}(x)\\ (n,P_y)=1}} 1.$$

This may be rewritten as

$$S^{(P)}(\mathscr{A}(x), y) = \sum_{n\in\mathscr{A}(x)} \sum_{d|(n,P_y)} \mu(d) = \sum_{d|P_y} \mu(d)|\mathscr{A}_d(x)|. \tag{3.1}$$

A straightforward substitution of (2.1) into (3.1) produces too many error terms arising out of the divisions of P_y when y is large and this becomes unwieldy. As is well known, Brun's idea was to consider χ_1, χ_2 satisfying $\chi_1(1) = \chi_2(1) = 1$ and

$$\sigma_1^{(2)}(n) = \sum_{d|n} \mu(d)\chi_2(d) \leqslant \sum_{d|n} \mu(d) \leqslant \sum_{d|n} \mu(d)\chi_1(d) = \sigma_1^{(1)}(n), \quad \forall n|P_y \tag{3.2}$$

so that when used in conjunction with (3.1) it would lead to bounds for $S^{(P)}(\mathscr{A}(x), y)$. Naturally the χ_i are to be chosen to keep the error terms under control.

We note that estimating $S_g(\mathscr{A}(x), y)$ can be considered as a general sieve problem for various reasons. Firstly, $S^{(P)}(\mathscr{A}(x), y)$ is a special case. Moreover

$$g_y(n) = \sum_{d|(n,P_y^*)} \mu(d)g^*(d),$$

where P^* is the set of all primes. So analogous to (3.1) we have

$$S_g(\mathscr{A}(x), y) = \sum_{n\in\mathscr{A}(x)} \sum_{d|(n,P_y^*)} \mu(d)g^*(d).$$

Just as in (3.2) suppose we had χ_1, χ_2, satisfying

$$(3.3)\qquad \sigma_{g^*}^{(2)}(n) = \sum_{d|n} \mu(d) g^*(d)\chi_2(d) \leqslant g(n) = \sum_{d|n}\mu(d)g^*(d) \leqslant \sum_{d|n}\mu(d)g^*(d)\chi_1(d) = \sigma_{g^*}^{(1)}(n), \qquad \forall n|P_y^*$$

then this would lead to the following bounds for $S_g(\mathscr{A}(x), y)$:

$$(3.4)\qquad X\sum_{d|P_y^*}\frac{\mu(d)g^*(d)\omega(d)\chi_2(d)}{d} - \sum_{d|P_y^*}|R_d(x)g^*(d)\chi_2(d)| \leqslant S_g(\mathscr{A}(x), y) \leqslant X\sum_{d|P_y^*}\frac{\mu(d)g^*(d)\omega(d)\chi_1(d)}{d} + \sum_{d|P_y^*}|R_d(x)g^*(d)\chi_1(d)|.$$

The idea is to choose $\chi_i(d) = 1$ on all 'small d' so as to get a main term of expected size and to make $\chi_i(d)$ small 'sufficiently often on large d' in order to control the remainder terms. There will be more on this in Sections 5 and 6.

The most interesting case of $S_g(\mathscr{A}(x), y)$ is when $y = x$. Apart from certain special situations the sieve does not yield good estimates for $S_g(\mathscr{A}(x), y)$ when y is close to x and so we cannot usually expect to estimate $S_g(\mathscr{A}(x))$ via $S_g(\mathscr{A}(x), y)$. However when $0 \leqslant g \leqslant 1$, we have $0 \leqslant g \leqslant g_y$ yielding $S_g(\mathscr{A}(x)) \leqslant S_g(\mathscr{A}(x), y)$, which is one reason we succeeded in applying this method to Probabilistic Number Theory; the second reason is (1.2). Sometimes it is possible to estimate $S_g(\mathscr{A}(x), y)$ asymptotically for all y using the sieve and hence for $y = x$ in particular; this feature is discussed toward the end of Section 6.

In many classical sieve problems the Buchstab identity

$$S^{(P)}(\mathscr{A}(x), y) = S^{(P)}(\mathscr{A}(x), y_1) + \sum_{\substack{y\leqslant p<y_1\\ p\in P}} S^{(P)}(\mathscr{A}_p(x), p)$$

has been used in treating large values y. One more reason for us to consider $S_g(\mathscr{A}(x), y)$ in a sieve context is because it also satisfies a similar identity, namely

$$(3.5)\qquad S_g(\mathscr{A}(x), y) = S_g(\mathscr{A}(x), y_1) + \sum_{y\leqslant p<y_1} g^*(p) S_g(\mathscr{A}_p(x), p).$$

To prove (3.5) first observe that

$$(3.6)\qquad S_g(\mathscr{A}(x), y) - S_g(\mathscr{A}(x), y_1) = \sum_{n\in\mathscr{A}(x)} (g_y(n) - g_{y_1}(n)).$$

Next,

$$(3.7)\qquad g_y(n) - g_{y_1}(n) = g_y(n)\{1 - \prod_{\substack{y\leqslant p<y_1\\ p|n}} g(p)\} = g_y(n)\{1 - \prod_{\substack{y\leqslant p<y_1\\ p|n}} (1-g^*(p))\}$$
$$= g_y(n)\{\sum_{\substack{p|n\\ y\leqslant p<y_1}} g^*(p) \prod_{\substack{q|n\\ y\leqslant q<p}} (1-g^*(q))\} = \sum_{\substack{p|n\\ y\leqslant p<y_1}} g^*(p) g_p(n).$$

So by putting together (3.6) and (3.7) we get

$$S_g(\mathscr{A}(x), y) - S_g(\mathscr{A}(x), y_1) = \sum_{y\leqslant p<y_1} g^*(p) \sum_{n\in\mathscr{A}_p(x)} g_p(n) = \sum_{y\leqslant p<y_1} g^*(p) S_g(\mathscr{A}_p(x), p)$$

which is (3.5). This Buchstab identity for $S_g(\mathscr{A}(x), y)$ will be useful in Sections 8 and 9.

Before proceeding further we need examples of χ_i satisfying (3.3) and for this the monotonicity principle discussed in the next section is useful.

4. A monotonicity principle.

THEOREM 1. *Let h and H be multiplicative functions such that $0 \leqslant h \leqslant H$ and $a(n)$ any arithmetical function. Then for all square-free n we have*

$$(4.1)\qquad \min_{\delta|n}\sum_{d|\delta} H(d)a(d) \leqslant \sum_{d|n} h(d)a(d) \leqslant \max_{\delta|n}\sum_{d|\delta} H(d)a(d).$$

In particular if

$$(4.2)\qquad \sum_{d|n} H(d)a(d) \geqslant 0 \quad \text{for all } n|P_y,$$

then

$$(4.3)\qquad \sum_{d|n} h(d)a(d) \geqslant 0 \quad \text{for all } n|P_y.$$

Proof. We prove only the first inequality in (4.1), the second being similar.

Let $n = p_1 \ldots p_r$ be square-free and $h(p_i) = x_i$, $i = 1, 2, \ldots, r$. We think of the x_i as variables in the interval $[0, H(p_i)]$. So

$$(4.4)\qquad \sum_{d|n} h(d)a(d) = L(x_1, \ldots, x_r)$$

can be considered as a linear form in each of the x_i. By taking successive minima over the x_i we obtain

$$(4.5)\qquad L(x_1, \ldots, x_r) \geqslant \min_{0\leqslant x_1\leqslant H(p_1)} L(x_1, x_2, \ldots, x_r) = L(e_1, x_2, \ldots, x_r)$$
$$\geqslant \min_{0\leqslant x_2\leqslant H(p_2)} L(e_1, x_2, x_3, \ldots, x_r) = L(e_1, e_2, x_3, \ldots, x_r)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\geqslant L(e_1, e_2, \ldots, e_r),$$

where each e_i is either 0 or $H(p_i)$. This means

$$(4.6)\qquad L(e_1, e_2, \ldots, e_r) = \sum_{d|\delta} H(d)a(d),$$

where δ is the product of all those p_i for which $e_i = H(p_i)$. Clearly (4.1) follows from (4.4), (4.5) and (4.6).

COROLLARY 1. *Suppose $H \geqslant 0$ is multiplicative and χ_1, χ_2 are such that for all $n|P_y$*

$$\sigma_H^{(2)}(n) = \sum_{d|n} \mu(d)H(d)\chi_2(d) \leqslant H^*(n)$$
$$= \sum_{d|n} \mu(d)H(d) \leqslant \sum_{d|n} \mu(d)H(d)\chi_1(d) = \sigma_H^{(1)}(n).$$

Then for any multiplicative h satisfying $0 \leqslant h \leqslant H$ we have for all $n|P_y$

$$\sigma_h^{(2)}(n) = \sum_{d|n} \mu(d)h(d)\chi_2(d) \leqslant h^*(n)$$
$$= \sum_{d|n} \mu(d)h(d) \leqslant \sum_{d|n} \mu(d)h(d)\chi_1(d) = \sigma_h^{(1)}(n).$$

Proof. The corollary follows from (4.2) and (4.3) by taking $a(d) = (-1)^{i-1}\mu(d)(\chi_i(d)-1)$ in Theorem 1, for $i = 1, 2$.

We think of (4.2) implying (4.3) (and in particular Corollary 1) as a 'monotonicity principle'. We had previously observed this principle in [1] with $H \equiv 1$ in Corollary 1 and proved it using Möbius inversion. We preferred the linear forms argument here since it works better when there are no size restrictions on H.

According to (4.1) of Theorem 1 the 'absolute errors' in Corollary 1 satisfy

$$|\sigma_h^{(i)}(n) - h^*(n)| \leqslant \max_{\delta|n} |\sigma_H^{(i)}(\delta) - H^*(\delta)|, \quad i = 1, 2.$$

For multiplicative functions bounded by 0 and 1 the following result shows that monotonicity is also exhibited in terms of the 'relative error' and for this Möbius inversion proves handy:

THEOREM 2. *For each square-free n consider the class C_n of all g such that $0 \leqslant g \leqslant 1$ and $g(n) \neq 0$. Suppose χ_1, χ_2 satisfy (3.2). Then the relative errors*

$$\left|\frac{\sigma_{g^*}^{(i)}(n)}{g(n)} - 1\right| \quad \text{for } i = 1, 2,$$

decrease with g^*, for $g \in C_n$.

Proof. By Möbius inversion we have

$$\sigma_{g^*}^{(i)}(n) = \sum_{d|n} g^*(d) \sum_{\delta|d} \mu\left(\frac{d}{\delta}\right)\sigma_1^{(i)}(\delta) = \sum_{\delta|n} \sigma_1^{(i)}(\delta)g^*(\delta)g\left(\frac{n}{\delta}\right). \tag{4.7}$$

Hence

$$\frac{\sigma_{g^*}^{(i)}(n)}{g(n)} - 1 = \sum_{1<\delta|n} \sigma_1^{(i)}(\delta)\frac{g^*(\delta)}{g(\delta)}. \tag{4.8}$$

From (3.2) we see that $\sigma_1^{(i)}(\delta)$ for $\delta > 1$ maintains the same sign. Also $g^*(\delta)/g(\delta) > 0$ and decreases with g^*. So Theorem 2 follows from (4.8).

Monotonicity is useful in providing functions χ_i satisfying (3.3) by starting with (3.2). We shall look into this closely in the next section.

5. Brun's sieve. To construct χ_i satisfying (3.2) it is customary to start with the decomposition

$$\sigma_1^{(i)}(n) = \sum_{d|\frac{n}{p}} \{\mu(d)\chi_i(d)+\mu(pd)\chi_i(pd)\} = \sum_{d|\frac{n}{p}} \mu(d)\{\chi_i(d)-\chi_i(pd)\} \tag{5.1}$$

in terms of a prime divisor p of the square-free integer $n > 1$. If the χ_i are such that for all $d|n/p$

$$(-1)^{i-1}\mu(d)\{\chi_i(d)-\chi_i(pd)\} \geqslant 0, \quad i = 1, 2, \tag{5.2}$$

then (3.2) follows from (5.1). Brun's original choice of χ_i is given by

$$\chi_i(n) = \begin{cases} 1 & \text{if } \nu(n) \leqslant 2s+i-1, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } i = 1, 2, \tag{5.3}$$

where s is a non-negative integer to be chosen optimally. Clearly (5.3) implies (5.2). The χ_i in (5.3) are the functions of 'Brun's pure sieve'. In this case

$$\sigma_1^{(i)}(n) = \binom{\nu(n)-1}{2s+i-1}(-1)^{i-1}, \quad \text{for } i = 1, 2. \tag{5.4}$$

It is however not necessary for (5.2) to hold for all $p|n$ because (3.2) will follow even if there is one prime p such that (5.2) holds for all $d|n/p$. It is convenient to take $p = q(n)$, the smallest prime divisor of $n > 1$, which is equivalent to saying $p < q(d)$ in (5.2) because of the convention $q(1) = \infty$. So, if (5.2) is to hold just for $p < q(d)$, then any χ_i satisfying

(i) $\chi_i(d) = 1$ or 0, $\forall\ d|P_y$,

(ii) $\chi_i(d) = 1 \Rightarrow \chi_i(t) = 1$, $\forall t|d$ (The χ_i are said to be divisor closed.),

(iii) $\chi_i(t) = 1$, $\mu(t) = (-1)^{i-1} \Rightarrow \chi_i(pt) = 1$, $\forall\ pt|P_y$, $p < q(t)$,

will suffice. These χ_i are the functions of the 'combinatorial sieve'. An example is Brun's refinement of the pure sieve obtained by partitioning $[2, y]$ into $2 = y_r < y_{r-1} < \ldots < y_1 = y_0 = y$ and letting

$$\chi_i(d) = \begin{cases} 1 & \text{if } \nu((d, P_{y_j,y})) \leqslant 2b-i+2j-1,\ j = 1, 2, \ldots, r, \\ 0 & \text{otherwise,} \end{cases} \tag{5.5}$$

where $P_{y_j,y} = P_y/P_{y_j}$. A detailed account of the combinatorial sieve with specific emphasis on Brun's choices is given in Chapter 2 of Halberstam and Richert [6] where proofs and discussion on many sieve results utilized here may be found.

Our interest in such χ_i lies in their applicability to (3.3). According to Corollary 1, if $0 \leqslant g^* \leqslant 1$, then (3.3) is a consequence of (3.2). In particular we have

THEOREM 3. *If χ_1, χ_2 are combinatorial sifting functions with $P = P^*$ and g^* satisfies $0 \leqslant g^* \leqslant 1$, then* (3.3) *holds for all $n|P_y^*$.*

Note that in Theorem 3 the specific properties of combinatorial sifting functions are used only to check (3.2) and these are not necessary to pass from (3.2) to (3.3) because the latter only requires monotonicity. It would however be interesting if special properties of the χ_i could be used to verify (3.3) for certain $g^* > 1$. An example is

THEOREM 4. *Let χ_1, χ_2 be as in* (5.3)*. Then* (3.3) *holds for all n if $0 \leqslant g^*(p) \leqslant 2$ for all p.*

Proof. In view of monotonicity it suffices to prove Theorem 4 for $g^*(n) = 2^{\nu(n)}$ in which case $g(n) = (-1)^{\nu(n)}$. We assume n is square-free and > 1.

With $\sigma_2^{(i)}(n)$ denoting $\sigma_{g*}^{(i)}(n)$ when $g^*(n) = 2^{\nu(n)}$ we have

$$\sigma_2^{(i)}(n) = \sum_{d|n} \mu(d)\chi_i(d) \sum_{\delta|d} 1 = \sum_{\delta|n} \sum_{e|\frac{n}{\delta}} \mu(\delta e)\chi_i(\delta e) \tag{5.6}$$

which by (5.3) is

$$\sum_{\delta|n} \mu(\delta) \sum_{\substack{e|\frac{n}{\delta} \\ \nu(\delta e) \leqslant r}} \mu(e) = \sum_{\delta|n} \mu(\delta) \sum_{\substack{e|\frac{n}{\delta} \\ \nu(e) \leqslant r-\nu(\delta)}} \mu(e), \tag{5.7}$$

where $r = 2s+i-1$. Next, because of (5.4)

$$(-1)^{i-1}\mu(\delta) \sum_{\substack{e|\frac{n}{\delta} \\ \nu(e) \leqslant r-\nu(\delta)}} \mu(e) = (-1)^{i-1-\nu(\delta)} \sum_{\substack{e|\frac{n}{\delta} \\ \nu(e) \leqslant r-\nu(\delta)}} \mu(e) \geqslant 0. \tag{5.8}$$

So from (5.6), (5.7) and (5.8) we have

$$(-1)^{i-1}\sigma_2^{(i)}(n) \geqslant 0. \tag{5.9}$$

But then notice that

$$\sigma_2^{(i)}(n) = 1 + \sum_{1<d|n} \mu(d)2^{\nu(d)}\chi_i(d) \equiv 1 \pmod 2. \tag{5.10}$$

Thus by (5.9) and (5.10) we deduce that

$$(-1)^{i-1}\sigma_2^{(i)}(n) \geqslant 1 \geqslant (-1)^{i-1}(-1)^{\nu(n)} = (-1)^{i-1}g(n) \tag{5.11}$$

as required.

In passing from (5.9) to (5.11) we made the convenient observation in (5.10) that $\sigma_2^{(i)}(n)$ was odd. This might not be the case if $g^*(p)$ were not even. Fortunately the monotonicity principle saved us from considering such g^*.

We could have proved Theorem 4 by decomposing the inner sum on the right-hand side of (5.6) as

$$\sum_{e|\frac{m}{p}} \{\mu(\delta e)\chi_i(\delta e) + \mu(\delta ep)\chi_i(\delta ep)\} = \sum_{e|\frac{m}{p}} \mu(\delta e)\{\chi_i(\delta e) - \chi_i(\delta ep)\}, \tag{5.12}$$

where $m = n/\delta$. This has the advantage of showing how special properties of Brun's pure sieve not shared by the more general combinatorial sieve could be exploited. Indeed since (5.2) holds in the case of the pure sieve for all $p|n$, the summands in (5.12) satisfy

$$(-1)^{i-1}\mu(\delta e)\{\chi_i(\delta e) - \chi_i(\delta ep)\} \geqslant 0. \tag{5.13}$$

This would not be guaranteed by the combinatorial sieve which satisfies (5.2) only for $p < q(d)$ whereas in (5.13), $p < q(\delta e)$ need not hold.

Theorem 4 it best possible in the sense that it fails if $g^*(p)$ is larger. For instance if $g^*(p) = 2+\varepsilon$ for all p, then $g(n) = (-1-\varepsilon)^{\nu(n)}$, whereas for fixed s, $\sigma_{g*}^{(i)}(n)$ behaves only like a polynomial in $\nu(n)$ of degree $2s+i-1$.

In view of the above remarks it would be of interest to determine for a given combinatorial sifting function χ_i, the set of all g^* for which (3.3) holds.

6. Sieve estimates for $S_g(\mathscr{A}(x), y)$. Throughout this section we assume that $0 \leqslant g^*(p) \leqslant 2$ for all p which is equivalent to $-1 \leqslant g \leqslant 1$.

An important parameter that arises naturally while estimating $S_g(\mathscr{A}(x), y)$ is sieve dimension $\varkappa = \varkappa_g$ which can be defined by

$$\sum_{p<y} \frac{\omega(p)g^*(p)}{p} \leqslant \varkappa \log\log y + c_1. \tag{6.1}$$

Clearly $\varkappa$ exists because $\omega(p) \ll 1$ and $0 \leqslant g^*(p) \leqslant 2$. It will also be convenient if there is $c_2 > 0$ such that

$$\left|1 - \frac{\omega(p)g^*(p)}{p}\right| \geqslant c_2 \quad \text{for all } p. \tag{6.2}$$

It is to be noted that in (6.2) we are not assuming $\omega(p)g^*(p) < p$.

Our objective is to replace

$$W_{g*}^{(i)}(y) = \sum_{d|P_y^*} \frac{\mu(d)g^*(d)\omega(d)\chi_i(d)}{d}$$

in (3.4) by

$$W_{g*}(y) = \sum_{d|P_y^*} \frac{\mu(d)g^*(d)\omega(d)}{d} = \prod_{p<y} \left(1 - \frac{\omega(p)g^*(p)}{p}\right).$$

Although it may not be the case that $\omega(p)g^*(p)/p < 1$, we certainly have $\omega(p)/p \leqslant 1$ for special sets. So with $H(d) = g^*(d)$ and $h(d) = \omega(d)g^*(d)/d$ we see from (3.3) and Corollary 1 that a direct replacement of $W_{g*}^{(i)}(y)$ by $W_{g*}(y)$ in (3.4) is not possible because the inequalities go the wrong way! Hence we need to take into account the error

$$\Theta_{g*}^{(i)}(y) = W_{g*}^{(i)}(y) - W_{g*}(y).$$

For this (4.7) is useful. To be more precise, with $g^*(d)$ and n in (4.7) replaced by $g^*(d)\omega(d)/d$ and P_y^* respectively, we have

$$\Theta_{g^*}^{(i)}(y) = W_{g^*}(y) \sum_{1<d|P_y^*} \sigma_1^{(i)}(d)G(d), \tag{6.3}$$

where

$$G(d) = \frac{\omega(d)g^*(d)}{d}\prod_{p|d}\left(1-\frac{\omega(p)g^*(p)}{p}\right)^{-1}. \tag{6.4}$$

We shall bound $\Theta_{g^*}^{(i)}(y)$ suitably using (6.3).

Suppose the χ_i are as in (5.3) and $r = 2s+i-1$. Then by (5.4)

$$\Big|\sum_{1<d|P_y^*} \sigma_1^{(i)}(d)G(d)\Big| \leqslant \sum_{1<d|P_y^*}\binom{\nu(d)}{r}G(d) \leqslant \sum_{m=r}^{\nu(P_y^*)}\binom{m}{r}\sum_{\substack{1<d|P_y^*\\ \nu(d)=m}} G(d) \tag{6.5}$$

$$\leqslant \sum_{m=r}^{\infty}\binom{m}{r}\frac{1}{m!}\Big(\sum_{p<y} G(p)\Big)^m$$

$$= \frac{1}{r!}\Big(\sum_{p<y} G(p)\Big)^r \exp\Big\{\sum_{p<y} G(p)\Big\}.$$

Note that (6.1), (6.2) and (6.4) yield

$$\sum_{p<y} G(p) \leqslant \frac{\varkappa}{c_2}\log\log y + O(1). \tag{6.6}$$

We put

$$r = \left[\frac{\varkappa \log\log y}{c_2\lambda}\right]+1, \tag{6.7}$$

where [ ] is the greatest integer function and λ which will satisfy

$$0 < \lambda e^{1+\lambda} < 1 \tag{6.8}$$

is to be chosen suitably. At any rate with r as in (6.7) we get

$$\frac{1}{r!}\Big(\sum_{p<y} G(p)\Big)^r \exp\Big\{\sum_{p<y} G(p)\Big\} \leqslant \left(\frac{e}{r}\right)^r (\lambda r)^r e^{\lambda r} = (\lambda e^{1+\lambda})^r$$

because of (6.6). This, when combined with (6.3) and (6.5), gives

$$|\Theta_{g^*}^{(i)}(y)| \leqslant |W_{g^*}(y)|(\lambda e^{1+\lambda})^r. \tag{6.9}$$

With regard to

$$E_{g^*}^{(i)}(y) = \sum_{d|P_y^*} |R_d(x)|g^*(d)\chi_i(d)$$

notice that if (2.4) holds then by (6.7)

$$E_{g^*}^{(i)}(y) \leqslant \Big(1+\sum_{p<y}\omega(p)g^*(p)\Big)^r \ll \exp\left\{\left(\frac{\varkappa\log\log y}{c_2\lambda}+1\right)\log y\right\}. \tag{6.10}$$

From (6.10) we can obtain an upper bound for $E_{g^*}^{(i)}(y)$ in the more general situation when (2.2) and (2.3) hold. To be precise

$$E_{g^*}^{(i)}(y) \leqslant \sum_{\substack{d|P_y^*\\ d\leqslant X^\beta/\log^V X}} |R_d(x)|2^{\nu(d)} + \sum_{\substack{d|P_y^*\\ d> X^\beta/\log^V X}}\left(\frac{X\log X}{d}+1\right)(2c_0)^{\nu(d)}\chi_i(d)$$

$$= \Sigma_1+\Sigma_2 \text{ respectively.}$$

Clearly

$$\Sigma_2 \leqslant X^{1-\beta}\log^{V+1}X\sum_{d|P_y^*}(2c_0)^{\nu(d)}\chi_i(d)$$

$$\leqslant \exp\left\{(1-\beta)\log X+(V+1)\log\log x+\left(\frac{\varkappa\log\log y}{c_2\lambda}+1\right)\log y\right\}$$

by comparison with (6.10). To bound Σ_1 we break it up into

$$\Sigma_1 = \sum_{\substack{d|P_y^*,\nu(d)\leqslant N\\ d\leqslant X^\beta/\log^V X}} + \sum_{\substack{d|P_y^*,\nu(d)> N\\ d\leqslant X^\beta/\log^V X}} = \Sigma_3+\Sigma_4 \text{ respectively,}$$

where $N = (U\log\log X)/2\log 2$. From (2.5) we quickly get

$$\Sigma_3 \ll_U \frac{X}{\log^U X}\cdot\log^{U/2}X = \frac{X}{\log^{U/2}X}.$$

As for Σ_4 note that by (2.2)

$$\Sigma_4 \ll X\log X\sum_{\substack{d|P_y^*\\ \nu(d)>N}}\frac{(2c_0)^{\nu(d)}}{d} \leqslant X\log X\sum_{k=N+1}^{\infty}(2c_0)^k\sum_{\substack{d|P_y^*\\ \nu(d)=k}}\frac{1}{d}$$

$$\leqslant X\log X\sum_{k=N+1}^{\infty}\frac{(2c_0\log\log y+c_3)^k}{k!} \ll_U X(\log X)^{-U\log U/4}.$$

Thus these bounds for Σ_j, $j = 1, 2, 3, 4$ imply that

$$E_{g^*}^{(i)}(y) \ll_U \frac{X}{\log^{U/2}X} \tag{6.11}$$

$$+\exp\left\{(1-\beta)\log X+(V+1)\log\log X+\left(\frac{\varkappa\log\log y}{c_2\lambda}+1\right)\log y\right\}.$$

It only remains to choose λ optimally to obtain an estimate for $S_g(\mathscr{A}(x), y)$ because

$$XW_{g^*}(y)-X|\Theta_{g^*}^{(2)}(y)|-E_{g^*}^{(2)}(y) \leqslant S_g(\mathscr{A}(x), y) \leqslant XW_{g^*}(y)+X|\Theta_{g^*}^{(1)}(y)|+E_{g^*}^{(1)}(y)$$

according to (3.4). On comparing (6.11) with (6.9) our choice is

$$\lambda = \frac{2\varkappa \log\log y \cdot \log y}{c_2 \beta \log X}. \tag{6.12}$$

Since (6.8) is a constraint on λ, naturally (6.12) imposes a restriction on y. We omit the details of the calculations and point out that with (6.12), (6.11), (6.10) and (6.9) we have established

THEOREM 5. *Let $\mathscr{A}$ be special and $-1 \leqslant g \leqslant 1$. Also let* (6.1) *and* (6.2) *hold. Then*

$$S_g(\mathscr{A}(x), y) = XW_{g*}(y)\{1+O_U(\log^{-U} X)\}$$

holds uniformly in g and for

$$y \leqslant X^{c_2\beta/2\varkappa\log\log X}. \tag{6.13}$$

If the stronger condition (2.4) *holds then*

$$S_g(\mathscr{A}(x), y) = XW_{g*}(y)\{1+O(e^{-c_4\alpha\log(\alpha/\varkappa)})\}+O(\sqrt{X})$$

holds uniformly in g and for y as in (6.13) *with $\beta = 1$.*

In order to estimate $S_g(\mathscr{A}(x), y)$ for y larger than that permitted by (6.13) we would need sifting functions superior to those of the pure sieve. But then we could not be sure that (3.3) holds for all g^* satisfying $0 \leqslant g^*(p) \leqslant 2$. Therefore suppose $0 \leqslant g^* \leqslant 1$ and that

$$1-\frac{\omega(p)}{p} > c_5 \quad \text{for all } p. \tag{6.14}$$

In addition we assume a stronger version of the dimension in equality (6.1), namely

$$\sum_{x_1<p<x_2} \frac{\omega(p)g^*(p)\log p}{p} \leqslant \varkappa_g \log\left(\frac{x_2}{x_1}\right)+c_6. \tag{6.15}$$

With these assumptions and in view of Theorems 2 and 3 all estimates which can be obtained for $S^{(P)}(\mathscr{A}(x), y)$ using combinatorial sifting functions hold also for $S_g(\mathscr{A}(x), y)$ with suitable modifications. To illustrate this we now state a result which is analogous to Theorem 2.5′ in [6], p. 82–83.

THEOREM 6. *Let $\mathscr{A}$ be special and $0 \leqslant g \leqslant 1$. Also let* (6.14) *and* (6.15) *hold. Then*

$$S_g(\mathscr{A}(x), y) = XW_{g*}(y)\{1+O(e^{-\alpha\beta\log(\alpha\beta/\varkappa)})+O_U(\log^{-U} X)\}$$

holds uniformly in y and g.

We shall omit the proof of this result which can be established using Brun's sieve (5.5) because it is similar to the one in Chapter 2 of [6] with $\omega(p)g^*(p)$ replacing $\omega(p)$. In fact since $0 \leqslant g^* \leqslant 1$, the simple inequality $|R_d(x)|g^*(d) \leqslant |R_d(x)|$ can be used in treating the remainder terms. So in proving Theorem 6 it is not necessary to consider terms like $|R_d(x)|c^{\nu(d)}$ with $c > 1$, as was done in deriving (6.11).

Theorem 6 yields an asymptotic estimate for $S_g(\mathscr{A}(x), y)$ when either $\alpha \to \infty$ with X, or $\varkappa \to 0$ as $X \to \infty$. When $\varkappa \to 0$ the asymptotic estimate holds for all $y \leqslant X$. In particular $S_g(x)$ can be asymptotically estimated as $x \to \infty$ if $\varkappa \to 0$. This has interesting implications in Probabilistic Number Theory (see § 8 of [3]).

For small α Theorem 2.1′ of [6], p. 65 provides explicit upper and lower bounds for $S^{(P)}(\mathscr{A}(x), y)$ and these are obtained using (5.5). Such bounds would now hold for $S_g(\mathscr{A}(x), y)$ when $0 \leqslant g \leqslant 1$, provided dimension is defined as in (6.15). In fact if better bounds for $S_g(\mathscr{A}(x), y)$ are required when $0 \leqslant g \leqslant 1$, then Rosser's sieve could be used because the Rosser functions are combinatorial in nature and Buchstab identities can be established for $S_g(\mathscr{A}(x), y)$ as in (3.5).

7. Exceptional cases. In order to apply a sieve method to estimate $S_g(\mathscr{A}(x), y)$ it is not necessary that conditions $0 \leqslant g^*(p) \leqslant 2$ (resp. $0 \leqslant g^*(p) \leqslant 1$) and (6.2) (resp. (6.14)) should hold for all p. The method would work as long as the exceptional set $P^{(0)}$ of primes, for which either of these conditions fail, is sparse. We now describe briefly how $S_g(\mathscr{A}(x), y)$ could be estimated when $P^{(0)}$ is finite and these arguments could be carried over in case $P^{(0)}$ is a sparsely distributed infinite collection of primes.

We begin with the decomposition

$$g(n) = h(n)\bar{h}(n),$$

where h, $\bar{h}$ are strongly multiplicative functions generated by

$$h(p) = \begin{cases} g(p), & p \in P^{(0)}, \\ 1, & p \notin P^{(0)}; \end{cases} \qquad \bar{h}(p) = \begin{cases} 1, & p \in P^{(0)}, \\ g(p), & p \notin P^{(0)}. \end{cases} \tag{7.1}$$

Clearly $g_y(n) = h_y(n)\bar{h}_y(n)$ and so

$$S_g(\mathscr{A}(x), y) = \sum_{n\in\mathscr{A}(x)} \bar{h}_y(n) \sum_{d|(n,P_y^*)} \mu(d)h_y^*(d) = \sum_{d|P_y^{(0)}} \mu(d)h_y^*(d) \sum_{n\in\mathscr{A}_d(x)} \bar{h}_y(n) \tag{7.2}$$

because according to (7.1), $h_y^*(d) = 0$ if $d \nmid P_y^{(0)}$. Since $P^{(0)}$ is finite, the divisors of $P_y^{(0)}$ are bounded. A convenient property of special sets $\mathscr{A}$ is that $\mathscr{A}_d$ is also special provided d is 'small'. In particular we are guaranteed that $\mathscr{A}_d$ is special in (7.2) because d is bounded. So if $\bar{h}_y$ satisfies the conditions on g^* in Theorems 5 and 6 then the sieve method would lead to the expression

$$S_g(\mathscr{A}(x), y) = \sum_{d|P_y^{(0)}} \mu(d)h_y^*\left\{\frac{X\omega(d)}{d} \prod_{p|\bar{P}_y^{(0)}} \left(1-\frac{\bar{h}_y^*(p)\omega(p)}{p}\right)(1+O(\eta_d(X, y)))\right\} \tag{7.3}$$

$$= XW_{g^*}(y) + O\Big(X|W_{\bar h^*}(y)|\{\max_{d|\bar P_y^{(0)}} \eta_d(X, y)\} \sum_{d|\bar P_y^{(0)}} \frac{|h_y^*(d)|\omega(d)}{d}\Big),$$

where $\bar P^{(0)} = P^* - P^{(0)}$ and $\eta_d(X, y)$ is any function which satisfies

(7.4) $$\Big|S_{\bar h}(\mathscr{A}_d(x), y) - \frac{X\omega(d)}{d}W_{\bar h^*}(y)\Big| \leqslant \frac{X\omega(d)}{d}W_{\bar h^*}(y)\eta_d(X, y).$$

From Theorems 5 and 6 it follows that we could choose η_d with the property

$$\max_{d|P_y^{(0)}} \eta_d(X, y) \ll \eta_1(X, y).$$

So from (7.3) and (7.4) we obtain

THEOREM 7. *Let $\mathscr{A}$ be special and $P^{(0)}$ finite.*

(a) *For $p \notin P^{(0)}$ suppose that $-1 \leqslant g(p) \leqslant 1$ and (6.2) hold. Then for $y \leqslant X^{c_2\beta/2\varkappa\log\log X}$*

(7.5) $$S_g(\mathscr{A}(x), y) = XW_{g^*}(y) + O(X|W_{\bar h^*}(y)|\eta_1(X, y)).$$

(b) *For $p \notin P^{(0)}$ suppose that $0 \leqslant g(p) \leqslant 1$ and (6.14) hold. Then (7.5) holds for $y \leqslant X$.*

COROLLARY 2. *In addition to the hypotheses of Theorem 7 (a) (resp. (b)), suppose that $\omega(p_0)g^*(p_0) = p_0$ for some $p_0 \in P^{(0)}$. Then for $p_0 < y \leqslant X^{c_2\beta/2\varkappa\log\log X}$ (resp. $p_0 < y \leqslant X$)*

$$|S_g(\mathscr{A}(x), y)| \ll X|W_{\bar h^*}(y)|\eta_1(X, y).$$

If $\omega(p_0)g^*(p_0) = p_0$, then $W_{g^*}(y) = 0$ for $y > p_0$ and so the 'main term' in Theorem 7 collapses. Such a phenomenon occurs when $g(n) = (-1)^{\nu(n)}$ and $\mathscr{A} = N$ because $\omega(2)g^*(2) = 2$. Denoting the sum $S_g(x, y)$ by $S_{-1}(x, y)$ in this case we obtain as a consequence of Corollary 2 and Theorem 5

(7.6) $$|S_{-1}(x, y)| \ll \frac{x}{\log^2 y}e^{-c_4\alpha\log(\alpha/2)} + \sqrt{x}, \quad \text{for } y \leqslant x^{c_2/4\log\log x}.$$

According to the remarks at the end of Section 8, this bound is not far away from the truth.

8. The sum $S_g(x, y)$. The main advantage in using the sieve on $S_g(\mathscr{A}(x), y)$ lies in the wide class of sets $\mathscr{A}$ which could be handled. But this put restrictions on the values $g(p)$. On the other hand if we concentrate only on $\mathscr{A} = N$ then other methods could be employed to treat $S_g(x, y)$ even when g takes arbitrarily large complex values.

For instance we could use the Perron integral. If $\mathrm{Re}(s) = \sigma > 1$ then

(8.1) $$G_y(s) = \sum_{n=1}^{\infty} \frac{g_y(n)}{n^s} = \prod_{p<y}\Big(1 + \frac{g(p)}{p^s - 1}\Big)\cdot\prod_{p \geqslant y}\Big(1 - \frac{1}{p^s}\Big)^{-1}$$
$$= \zeta(s)\prod_{p<y}\Big(1 - \frac{g^*(p)}{p^s}\Big)$$

is analytic in s provided $|g(p)|$ is not too large and (8.1) gives an analytic continuation of $G_y(s)$ into $\sigma \leqslant 1$. By the Perron formula

(8.2) $S_g(x, y)$
$$= \frac{1}{2\pi i}\int_{a-iT}^{a+iT} \frac{x^s G_y(s)ds}{s} + O\Big(\sum_{n=1}^{\infty} \frac{|g_y(n)x^a|}{n^a}\min\Big\{\frac{1}{T\left|\log\frac{x}{n}\right|}, \log\frac{1}{T\left|\log\frac{x}{n}\right|}\Big\}\Big)$$

where $a > 1$ and $T > 2$ are to be chosen suitably.

For the sum

$$\Phi(x, y) = \sum_{n \leqslant x, q(n) \geqslant y} 1$$

which is a particular case of $S_g(x, y)$, de Bruijn [4] used (8.2) to obtain estimates when α is large. The arguments in [4] could be applied to $S_g(x, y)$ with some minor changes. If g is complex valued and $|g(p)| \leqslant C$, then the inequalities

$$\sum_{n \leqslant x} \frac{|g(n)|}{n} < \prod_{p<x}\Big(1 + \frac{C}{p-1}\Big) \ll \log^C x$$

and

$$\Big|\prod_{p<y}\Big(1 - \frac{g^*(p)}{p^s}\Big)\Big| \leqslant \prod_{p<y}\Big(1 + \frac{C+1}{p^\sigma}\Big)$$

will be useful in dealing with the error terms in (8.1) and in estimating the contribution from the integral by closing the contour around the pole at $s = 1$. We omit the details which are similar to those in [4] and state

THEOREM 8. *Let g be complex valued and $|g(p)| \leqslant C$ for all p. Then*

$$S_g(x, y) = x\prod_{p<y}\Big(1 - \frac{g^*(p)}{p}\Big) + O(x\log^{C+2} y,\ e^{-\alpha\log\alpha - \alpha\log\log\alpha + O(\alpha)}).$$

This result is useful mainly when α is large and in fact the estimate becomes trivial if α is small.

If $g(p)$ is well behaved, then, for small α, a better approach to $S_g(x, y)$ is via the Buchstab identity

(8.3) $$S_g(x, y) = S_g(x) + \sum_{y \leqslant p < x} g^*(p)S_g(x/p, p)$$

which is a special case of (3.5). Indeed if $g(p)$ has an average value then $S_g(x)$ can be estimated using Perron's formula and the Riemann zeta function. This serves as a starting point in (8.3) to evaluate $S_g(x, y)$ by induction on $[\alpha]$ because $\log(x/p)/\log p \leqslant \alpha - 1$. Buchstab first used this idea to evaluate $\Phi(x, y)$ for an arbitrary but fixed α. Subsequently de Bruijn [4] improved Buchstab's approach and estimated $\Phi(x, y)$ uniformly in α. We note that de Bruijn's ideas could be applied more generally to $S_g(x, y)$ when $g(p)$ for $2 \leqslant p \leqslant x$ has an average value which is nearly constant; that is the average differs from a complex number z by an amount which shrinks to zero sufficiently fast as $x\to\infty$. In particular the method shows that for $3<\alpha\leqslant\sqrt{\log y}$

$$|S_{-1}(x,y)|\ll\frac{x}{\log^3 y}e^{-\alpha\log\alpha-\alpha\log\log\alpha+O(\alpha)} \tag{8.4}$$

and that this bound is best possible except for the implicit constants. From (8.4) and (7.6) we see that Corollary 2 cannot be improved substantially.

We plan to discuss applications of de Bruijn's method to $S_g(x,y)$ in a subsequent paper with particular emphasis on $S_{-1}(x,y)$.

9. The small sieve. Erdős and Ruzsa [5] have established the following 'small sieve' result: There exists an absolute constant c such that if $x>1$ and $K>0$ are arbitrary and P any set of primes for which

$$\sum_{p\in P}\frac{1}{p}\leqslant K,$$

then

$$S^{(P)}(x)=\sum_{\substack{n\leqslant x\\(n,P(x))=1}}1\geqslant e^{-e^{cK}}x. \tag{9.1}$$

The interest in (9.1) lies in its uniformity with respect to P. Apart from the value of c inequality (9.1) is best possible as can be seen by taking $S^{(P)}(x)$ to be

$$\psi(x,y)=\sum_{\substack{n\leqslant x\\p|n,p\leqslant y}}1,\qquad K\sim\log\alpha.$$

We have obtained a multiplicative generalization of the small sieve result, namely

THEOREM 9. *There exists an absolute constant c such that if $x>1$, $K>0$ and g satisfies $0\leqslant g\leqslant 1$ and*

$$\sum_{p\leqslant x}\frac{g^*(p)}{p}\leqslant K, \tag{9.2}$$

then uniformly in x, K and g we have

$$S_g(x)\geqslant e^{-e^{cK}}x.$$

Proof. We shall utilize the method of Erdős–Ruzsa suitably generalized for our purpose. Throughout this section we assume that $0\leqslant g,\ g^*\leqslant 1$.

At the outset we observe that

$$\sum_{n\leqslant x}\frac{g(n)}{n}\geqslant\log x\prod_{p\leqslant x}\left(1-\frac{g^*(p)}{p}\right). \tag{9.3}$$

To realize (9.3) note that

$$\log x\leqslant\sum_{n\leqslant x}\frac{\mu^2(n)}{n}=\sum_{n\leqslant x}\frac{\mu^2(n)}{n}\sum_{d|n}g(d)g^*\left(\frac{n}{d}\right)\leqslant\left(\sum_{d\leqslant x}\frac{g(d)}{d}\right)\left(\sum_{d\leqslant x}\frac{\mu^2(d)g^*(d)}{d}\right)$$

and so

$$\sum_{d\leqslant x}\frac{g(d)}{d}\geqslant\log x\left(\sum_{d\leqslant x}\frac{\mu^2(d)g^*(d)}{d}\right)^{-1}$$
$$\geqslant\log x\prod_{p\leqslant x}\left(1+\frac{g^*(p)}{p}\right)^{-1}\geqslant\log x\prod_{p\leqslant x}\left(1-\frac{g^*(p)}{p}\right)$$

as claimed.

Next let $y=x^{1-\delta}$, $0<\delta<1/2$. We claim that

$$S_g(x,y)\geqslant\delta e^{-K}x. \tag{9.4}$$

To prove (9.4) consider an integer $b\leqslant x^{\delta}$ and $p>x^{1-\delta}$. If $m=pb\leqslant x$, then $g_y(m)=g(b)$. Given b, there are at least $\pi(x/b)-\pi(x^{1-\delta})$ such m. So by (9.3) and (9.2)

$$\sum_{n\leqslant x}g_y(n)\geqslant\sum_{b\leqslant x^{\delta}/2}g(b)\{\pi(x/b)-\pi(x^{1-\delta})\}\gg\frac{x}{\log x}\sum_{b\leqslant x^{\delta}/2}\frac{g^*(b)}{b}$$
$$\gg\delta x\prod_{p\leqslant x^{\delta}/2}\left(1-\frac{g^*(p)}{p}\right)\gg\delta e^{-K}x$$

as claimed.

We now attempt to evaluate $S_g(x)$ from (9.4) by appeal to (8.3). More precisely let $y=x^{1-(1/k)}$, where $k=e^{K+2}$. Then

$$S_g(x)=S_g(x,y)-\sum_{y\leqslant p<x}g^*(p)S_g\left(\frac{x}{p},p\right)\gg\frac{e^{-K}}{k}x-x\sum_{y\leqslant p<x}\frac{g^*(p)}{p} \tag{9.5}$$
$$\geqslant(c_7e^{-2K}-\lambda)x,$$

where

$$\lambda=\sum_{y\leqslant p<x}\frac{g^*(p)}{p}. \tag{9.6}$$

If $\lambda<c_7e^{-2K}/2$, then (9.5) gives a lower bound for $S_g(x)$ of the form $c_8e^{-2K}x$ which is better than Theorem 9. Therefore we shall assume from now on that $\lambda>c_8e^{-2K}$.

For convenience let

$$\gamma(K)=\inf\left\{\frac{S_g(x)}{x}\ \middle|\ g^*\ \text{satisfying (9.2)}\right\}.$$

We shall bound $\gamma(K)$ from below by real type induction on K. In doing so we may assume $x > e^{e^{cK}}$ with c sufficiently large because otherwise Theorem 9 is trivial.

To start the induction we use Theorem 3 with χ_2 given by $s = 0$ in (5.3). So

(9.7)
$$S_g(x) = \sum_{n\leqslant x}\sum_{d|n}\mu(d)g^*(d) \geqslant \sum_{n\leqslant x}\{1-\sum_{p|n}g^*(p)\} \geqslant x - x\sum_{p\leqslant x}\frac{g^*(p)}{p} \geqslant (1-K)x.$$

Hence for $K \leqslant 1/2$ Theorem 9 is true.

Next, let Q denote the set of integers $m \leqslant x$ which have a prime divisor $q > x^{1/k}$. Let $m = qb$. Note that $g(m) \geqslant g(q)g(b)$. Also m has at most k representations of the form qb since it has at most k such prime divisors q. Therefore

$$S_g(x) \geqslant \sum_{m\in Q} g(m) \geqslant \frac{1}{k}\sum_{x^{1/k}\leqslant q\leqslant x} g(q) \sum_{b\leqslant x/q} g(b). \tag{9.8}$$

The integers b in (9.8) can only have prime divisors $p < x^{1-(1/k)}$. So by real type induction and (9.6) we have

$$\begin{aligned}(9.9)\quad S_g(x) &\geqslant \frac{\gamma(K-\lambda)}{k}x\sum_{x^{1/k}\leqslant q\leqslant x}\frac{g(q)}{q} \geqslant \frac{\gamma(K-\lambda)x}{k}\Big(\sum_{x^{1/k}\leqslant p\leqslant x}\frac{1}{p} - \sum_{p\leqslant x}\frac{g^*(p)}{p}\Big)\\ &\geqslant \frac{\gamma(K-\lambda)x}{k}\Big(\log k - K + O\Big(\frac{1}{\log x}\Big)\Big) \geqslant x\gamma(K-\lambda)e^{-(K+2)}.\end{aligned}$$

From (9.7) and (9.9) we see that $\gamma(K)$ is bounded below by any $\gamma_0(K)$ which satisfies

$$\gamma_0(K) < \begin{cases}(1-K) & \text{for } K \leqslant 1/2,\\ e^{-(K+2)}\gamma_0(K-c_8e^{-2K}) & \text{for } K > 1/2.\end{cases}$$

Clearly $\gamma_0(K) = e^{-e^{c_9K}}$ with a large c_9 works and that proves Theorem 9.

Regarding an upper bound for $S_g(x)$ which is uniform in $0 \leqslant g \leqslant 1$, the answer is contained in Theorem 6. Indeed since $g(n) \leqslant g_y(n)$ we have shown more generally for special sets that

$$S_g(\mathscr{A}(x)) \ll X\prod_{p\leqslant X}\Big(1-\frac{g^*(p)\omega(p)}{p}\Big) \ll X\exp\Big\{-\sum_{p\leqslant X}\frac{g^*(p)\omega(p)}{p}\Big\} \tag{9.10}$$

and this is best possible except for the value of the implicit constant.

Instead of just a lower bound for $S_g(x)$ we desire, in the spirit of (9.10), a lower bound for $S_g(\mathscr{A}(x))$ given that

$$\sum_{p\leqslant x}\frac{g^*(p)\omega(p)}{p} \leqslant K.$$

Unless $\mathscr{A}$ is rather restrictive this problem involves great difficulties. One instance where an analogue of Theorem 9 could be proved is by taking $\mathscr{A}$ to be

integers n for which $\nu(n)$ is odd (or even). However, even for $\mathscr{A}$ as in (E−1) of Section 2 we are unable to prove a result like Theorem 9; in fact such an inequality would imply a deep conjecture of Erdős on the distribution of numbers of the type $p+a$ devoid of large prime factors.

10. Concluding remarks. For the classical sieve problem of estimating $S^{(P)}(\mathscr{A}(x), y)$ the dimension $\varkappa$ measures the average amount of sieving done per prime p. When $0 \leqslant g \leqslant 1$ the quantity $S_g(\mathscr{A}(x), y)$ can be interpreted as the residual amount after $\mathscr{A}(x)$ has been sifted through $p < y$ because the weights corresponding to $n \equiv 0 \pmod p$ shrink by a factor $g(p)$ when sieving with p. But when $0 \leqslant g \leqslant 1$ does not hold, the sum $S_g(\mathscr{A}(x), y)$ cannot be given a sieve interpretation in the classical sense. However, it is interesting that as long as $-1 \leqslant g \leqslant 1$, sieve methods could be used to estimate $S_g(\mathscr{A}(x), y)$ despite the oscillation in sign of the summands $g_y(n)$. In this case estimating $S_g(\mathscr{A}(x), y)$ is like treating $S^{(P)}(\mathscr{A}(x), y)$ in dimension $\varkappa_g$. In particular if $S^{(P)}(\mathscr{A}(x), y)$ has dimension $\varkappa$ then $S_g(\mathscr{A}(x), y)$ with $g(n) = (-1)^{\nu(n)}$ is like performing a $2\varkappa$-dimensional sieve on $\mathscr{A}(x)$! It was this unusual feature that motivated this paper.

We have made many observations about various questions which arise naturally in connection with results established here. We plan to discuss some of these problems in detail on a later occasion.

References

[1] K. Alladi, *A study of the moments of additive functions using Laplace transforms and sieve methods*, Proc. 4th MATSCIENCE Conf. Number Theory, Ootacamund, India (1984), Springer Lect. Notes 1122, pp. 1–37.

[2] — *Moments of additive functions and the sequence of shifted primes*, Pacific J. Math. 118 (1985), 261–275.

[3] — *A new application of the sieve to Probabilistic Number Theory*, Proc. Conf. Number Th., Austin, Texas (1982), Univ. Texas Press, pp. 1–27.

[4] N. G. de Bruijn, *On the number of uncancelled elements in the sieve of Eratosthenes*, Indag. Math. 12 (1950), 247–256.

[5] P. Erdős and I. Ruzsa, *On the small sieve. I. Sifting by primes*, J. Number Theory 12 (1980), 385–394.

[6] H. Halberstam and H. -E. Richert, *Sieve Methods*, Academic Press, N. Y., 1974.

THE INSTITUTE OF MATHEMATICAL SCIENCES
Madras 600 113, India

Current address:
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF FLORIDA
Gainesville, Florida
32611, U.S.A.

Received on 18.3.1986 (1607)