

# LINNIK'S THEOREM ON GOLDBACH NUMBERS IN SHORT INTERVALS

by D. A. GOLDSTON†

(Received 11 March, 1989)

**1. Introduction.** The Goldbach conjecture states that every even number larger than 2 can be written as the sum of two primes. We shall therefore call an even number a Goldbach number if it can be written as the sum of two primes in at least one way. It has been known for a long time that almost all even numbers are Goldbach numbers. In fact, Montgomery and Vaughan [14] have shown that if  $E(N)$  denotes the number of even numbers less than or equal to  $N$  which are not Goldbach numbers, then there exists an absolute constant  $\delta > 0$  such that

$$E(N) \ll N^{1-\delta}. \tag{1}$$

Another question concerning Goldbach numbers is to ask whether they always exist in short intervals, or equivalently whether there exist long sequences of consecutive even numbers which are not Goldbach numbers. The first result in this direction is due to Linnik [12], who proved, assuming the Riemann Hypothesis, that for any  $\epsilon > 0$  and  $N$  sufficiently large, the interval  $[N, N + (\log N)^{3+\epsilon}]$  contains a Goldbach number. This result was improved by Kátai [11], and later and independently by Montgomery and Vaughan [14]; they showed that assuming the Riemann Hypothesis there exists an absolute constant  $C$  such that for all sufficiently large  $N$  the interval  $[N, N + C(\log N)^2]$  contains a Goldbach number. Montgomery and Vaughan also proved unconditionally that the interval  $[N, N + N^{7/72+\epsilon}]$  will contain Goldbach numbers for  $N$  sufficiently large.

The method of Kátai and Montgomery-Vaughan uses a simple argument to relate Goldbach numbers to primes in short intervals, and in particular to a result of Selberg. Define

$$J(N, h) = \int_1^N (\vartheta(x+h) - \vartheta(x) - h)^2 dx, \tag{2}$$

where  $\vartheta(x) = \sum_{p \leq x} \log p$ , and  $p$  will always denote a prime number. Now suppose that the interval  $[N, N+h]$  contains no Goldbach numbers. Then at least one of the two intervals  $[y, y + \frac{1}{2}h]$  or  $[N-y, N-y + \frac{1}{2}h]$  contains no primes. Next cover the interval  $[1, N]$  by the set of intervals  $[N/2 + kh/2, N/2 + (k+1)h/2]$ , where  $k$  is an integer, and  $-N/2h < k < N/2h$ . Of these  $N/h + O(1)$  intervals at least  $N/2h - O(1)$  contain no primes, and hence the set  $E = \{y : 1 \leq y \leq N, [y, y+h/4] \text{ contains no primes}\}$  has measure  $\gg N$ . We conclude

$$J(N, h/8) \geq \int_E (\vartheta(x+h/8) - \vartheta(x) - h/8)^2 dx = \int_E (h/8)^2 dx \gg h^2 N. \tag{3}$$

Selberg [17] proved that assuming the Riemann Hypothesis,

$$J(N, h) \ll hN \log^2 N, \tag{4}$$

† Research supported in part by NSF Grant DMS8705710

which implies that (3) can hold only if  $h^2N \ll hN \log^2 N$  and hence  $h \ll \log^2 N$ . The unconditional result is proved by a similar argument with an unconditional bound for  $J(N, h)$ , but also uses Huxley’s result on primes in intervals  $[N, N + N^{7/12+\epsilon}]$ .

The purpose of this paper is to show how Linnik’s argument can be modified to give the results of Kátai and Montgomery-Vaughan, and also give asymptotic formulas for the number of representations of even numbers as sums of two primes in short intervals. To state our results, let

$$r(n) = \sum_{p+p'=n} 1, \quad R(n) = \sum_{p+p'=n} \log p \log p'. \tag{5}$$

**THEOREM.** *For any integer  $1 \leq K \leq N/\log^3 N$ , we have, for all sufficiently large  $N$ ,*

$$\left| \sum_{n=-K}^K (K - |n|)R(N + n) - K^2N \right| \leq J(N, K) + O(KN) + o(K^2N). \tag{6}$$

**COROLLARY 1.** *Suppose that  $J(N, K) = o(K^2N)$ . Then*

$$\sum_{n=N}^{N+K} r(n) \sim K \frac{N}{\log^2 N} \quad \text{as } N \rightarrow \infty. \tag{7}$$

The result in Corollary 1 shows that one obtains an asymptotic formula for the number of  $p + p'$  in the interval  $[N, N + K]$ . Since there is at most one way to write an odd number as a sum of two primes, we see that *on average* the number of representations of an even number  $n$  as a sum of two primes is  $\sim 2N/\log^2 N$ , where the average is over all such  $n$  in the interval  $[N, N + K]$ . If we ask how many of the even numbers in this interval are actually Goldbach numbers, we can not prove that almost all of them are Goldbach numbers (see Ramachandra [15] for results of this type.) However, by using a standard argument from the Schnirelman method, we can prove a positive proportion of the numbers in  $[N, N + K]$  are Goldbach numbers, provided  $K$  is not too small. This type of result was first obtained by Ramachandra [16].

**COROLLARY 2.** *Suppose that  $J(N, K) = o(K^2N)$  and  $K \geq C \log^2 N$ , where  $C$  is a sufficiently large absolute constant. Then for all sufficiently large  $N$ , we have*

$$\sum_{\substack{N \leq n \leq N+K \\ r(n) \geq 1}} 1 \gg K. \tag{8}$$

A. Ivić has found that the condition  $K \geq C \log^2 N$  in Corollary 2 can be weakened to  $K \geq C \log^{1+\epsilon} N$ , for any  $\epsilon > 0$ .

The integral  $J(N, h)$  has been studied by many mathematicians with the goal of finding the smallest function  $g(N)$  such that the estimate  $J(N, h) = o(h^2N)$  holds for all  $h \geq g(N)$ . Usually  $J(N, h)$  is defined with the function  $\psi$ , which counts primes and prime powers, in place of  $\vartheta$ , but the same estimates hold for  $J(N, h)$  in either case. Unconditionally it is known that  $g(N) = N^{1/6+\epsilon}$  is acceptable ([14], or [10, p. 319]). Assuming either the Lindelöf hypothesis [10] or the Elliott-Halberstam conjecture [8] gives  $g(N) = N^\epsilon$ . Assuming the Density Hypothesis in the form  $N(\sigma, T) \ll T^{2(1-\sigma)} \log^A T$  implies we may take  $h(N) = \log^B N$ , where  $B$  is a constant that depends on  $A$  [10]. Assuming the Riemann Hypothesis, Selberg’s result in equation (4) implies that  $J(N, h) = o(h^2N)$  for any  $h/\log^2 N \rightarrow \infty$ . The true behaviour of  $J(N, h)$  is quite likely

given by

$$J(N, h) \sim hN \log(N/h), \quad 1 \leq h \leq N^{1-\epsilon}, \tag{9}$$

which in [5] has been proved, assuming the Riemann Hypothesis, to be equivalent to a form of Montgomery’s pair correlation conjecture. Thus, subject to (9), we have that Corollary 1 holds for  $K/\log N \rightarrow \infty$ , and further the Theorem implies immediately that for all sufficiently large  $N$  every interval of the form  $[N, N + (2 + \epsilon)\log N]$  will contain Goldbach numbers. This is nearly the limit of our method, for in [4] it has been shown that, for  $1 \leq h \leq \log^A N$ , where  $A$  is any positive constant,

$$J(N, h) > (\frac{1}{2} - \epsilon)hN \log N, \tag{10}$$

which shows that the theorem cannot detect Goldbach numbers in intervals shorter than  $\log N$ .

An important aspect of the theorem is that to prove the existence of Goldbach numbers we only need to obtain the upper bound estimate  $J(N, h) < (1 - \epsilon)h^2N$ . One way to attempt to obtain this type of estimate is through sieve methods. Let

$$Z(n; k) = \sum_{\substack{p, p' \leq n \\ p' - p = k}} \log p' \log p. \tag{11}$$

It has been conjectured by Hardy and Littlewood [7] that there are asymptotic formulas for both the Goldbach problem and the “conjugate” prime pair problem. The conjecture is that

$$R(n) = \mathfrak{S}(n)n + o(n), \quad Z(n; k) = \mathfrak{S}(k)n + o(n), \quad \text{as } n \rightarrow \infty, \tag{12}$$

where for  $n$  odd  $\mathfrak{S}(n) = 0$ , and for even  $n$

$$\mathfrak{S}(n) = 2C \prod_{\substack{p \mid n \\ p \geq 3}} \left( \frac{p-1}{p-2} \right), \quad C = \prod_{p \geq 3} \left( 1 - \frac{1}{(p-1)^2} \right). \tag{13}$$

Sieve methods allow us to show both  $R(n)$  and  $Z(n; k)$  are bounded by a constant multiple of the conjectured factor in (12) (see [6]). In the case of  $Z(n; k)$ , we have, for any  $\epsilon > 0$ ,

$$Z(n; k) \leq (A + \epsilon)\mathfrak{S}(k)n. \tag{14}$$

Equation (14) holds for all  $k$  with  $A = 3.9171$  [1], and we may take  $A = 3.454$  if  $k$  is restricted to  $k \leq \log^D n$ , where  $D$  is any positive number [2]. If we could obtain (14) with  $A < 2$ , it would have the following consequence.

**COROLLARY 3.** *If (14) holds with  $A < 2$ , then for all sufficiently large  $N$  the interval  $\left[ N, N + \frac{1}{1 - A/2 - \epsilon} \log N \right]$  contains Goldbach numbers.*

In the last section we discuss a conjecture on  $J(N, h)$  which is suggested by a strengthening of (12). This conjecture may be connected heuristically by the method of [5] with the number of simple zeros of the Riemann zeta-function, and is consistent with the expectation that all the zeros are simple.

I would like to thank A. Ivić for his helpful comments on this paper.

**2. The Main Lemma.** We begin with the additive generating function for primes

$$S(\alpha) = \sum_{p \leq N} \log p e(p\alpha); \quad e(u) = e^{2\pi i u}. \tag{15}$$

Define

$$r(N, n) = \sum_{\substack{p, p' \leq N \\ p+p'=n}} 1, \quad R(N, n) = \sum_{\substack{p, p' \leq N \\ p+p'=n}} \log p \log p'. \tag{16}$$

Thus  $r(n) = r(N, n)$  for  $n \leq N$ , and  $R(n) = R(N, n)$  for  $n \leq N$ . Notice  $R(N, n)$  is the  $n$ th Fourier coefficient of  $S(\alpha)^2$ . Our starting point is the obvious relation

$$R(N, n) = \int_0^1 S(\alpha)^2 e(-n\alpha) d\alpha = \int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha + \int_{\beta}^{1-\beta} S(\alpha)^2 e(-n\alpha) d\alpha. \tag{17}$$

Linnik’s idea was that the first integral above can be evaluated if  $\beta$  is small, and on average the second integral is smaller than the first integral. We evaluate the first integral in the following lemma.

LEMMA 1. Let  $N \geq 2$ ,  $1/N \leq \beta \leq 1/2$ , and  $0 \leq n \leq 2N$ . Then we have

$$\begin{aligned} \int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha &= \min(n - 1, 2N - n + 1) + O(\beta^2 J(N, 1/2\beta)) \\ &\quad + O(\sqrt{(N\beta^2 J(N, 1/2\beta))}) + O(\beta N) + O(\beta^{-1/2} N^{1/2}). \end{aligned} \tag{18}$$

The same result holds with  $S(\alpha)^2$  replaced with  $|S(\alpha)|^2$  if the first term on the right of (18) is replaced by  $\max(N - n, 0)$ .

*Proof.* One form of the prime number theorem states that

$$S(0) = \sum_{p \leq N} \log p \sim N \quad \text{as } N \rightarrow \infty,$$

which suggests that

$$I(\alpha) = \sum_{k=1}^N e(k\alpha) \tag{19}$$

is an approximation for  $S(\alpha)$  when  $\alpha$  is close to zero. Thus

$$\begin{aligned} \int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha &= \int_{-\beta}^{\beta} (I(\alpha) + S(\alpha) - I(\alpha))^2 e(-n\alpha) d\alpha \\ &= \int_{-\beta}^{\beta} I(\alpha)^2 e(-n\alpha) d\alpha + O\left(\int_{-\beta}^{\beta} |I(\alpha)| |S(\alpha) - I(\alpha)| d\alpha\right) \\ &\quad + O\left(\int_{-\beta}^{\beta} |S(\alpha) - I(\alpha)|^2 d\alpha\right) \\ &= I_1 + O(I_2) + O(I_3). \end{aligned} \tag{20}$$

Now

$$\begin{aligned} I_1 &= \int_{-1/2}^{1/2} I(\alpha)^2 e(-n\alpha) d\alpha + O\left(\int_{\beta}^{1/2} \frac{1}{\alpha^2} d\alpha\right) \\ &= \sum_{\substack{1 \leq k, k' \leq N \\ k+k'=n}} 1 + O(1/\beta) \\ &= \min(n - 1, 2N - n + 1) + O(1/\beta). \end{aligned}$$

Next, we estimate  $I_3$  using a lemma of Gallagher ([3], [13]), which states that for  $T\delta = \frac{1}{2}$ , and  $\sum_{\mu} |c(\mu)| < \infty$ ,

$$\int_{-T}^T \left| \sum_{\mu} c(\mu)e(\mu\alpha) \right|^2 d\alpha \ll \int_{-\infty}^{\infty} \left| \frac{1}{\delta} \sum_{|\mu-x| < \delta/2} c(\mu) \right|^2 dx.$$

Hence, on letting  $\lambda(k) = \log p$  if  $k = p$ , and  $\lambda(k) = 0$  otherwise, we have

$$\begin{aligned} I_3 &= \int_{-\beta}^{\beta} \left| \sum_{k=1}^N (\lambda(k) - 1)e(k\alpha) \right|^2 d\alpha \\ &\ll \int_{-\infty}^{\infty} \left| \beta \sum_{\substack{1 \leq k \leq N \\ |k-x| < 1/4\beta}} (\lambda(k) - 1) \right|^2 dx \\ &\ll \beta^2 \int_{1-1/2\beta}^1 (\vartheta(x + 1/2\beta) - (x + 1/2\beta) + O(1))^2 dx \\ &\quad + \beta^2 \int_1^N (\vartheta(x + 1/2\beta) - \vartheta(x) - 1/2\beta + O(1))^2 dx. \end{aligned}$$

The first integral in the last line is by the Chebyshev estimate  $\vartheta(x) \ll x$  bounded by  $O(1/\beta)$ . Since  $(a + b)^2 \leq 2(a^2 + b^2)$ , we conclude

$$I_3 \ll 1/\beta + \beta^2 J(N, 1/2\beta) + \beta^2 N.$$

We estimate  $I_2$  by the Cauchy-Schwarz inequality and the estimates obtained for  $I_1$  and  $I_3$ . We thus see, for  $1/N \leq \beta \leq 1/2$ ,

$$\begin{aligned} I_2 &\ll (N + 1/\beta)^{1/2} (\beta^2 \{N + J(N, 1/2\beta) + 1/\beta^3\})^{1/2} \\ &\ll \beta N + \sqrt{(\beta^2 N J(N, 1/2\beta))} + \beta^{-1/2} N^{1/2}. \end{aligned}$$

The first part of the lemma now follows from these bounds and (20) on noting that for  $\beta \geq 1/N$ ,  $1/\beta \ll \beta^{-1/2} N^{1/2}$ .

The only change in replacing  $S(\alpha)^2$  by  $|S(\alpha)|^2$  is that  $I_1$  is replaced by

$$\begin{aligned} \int_{-1/2}^{1/2} |I(\alpha)|^2 e(-n\alpha) d\alpha + O\left(\int_{\beta}^{1/2} \frac{1}{\alpha^2} d\alpha\right) &= \sum_{\substack{1 \leq k, k' \leq N \\ k' - k = n}} 1 + O(1/\beta) \\ &= \max(N - n, 0) + O(1/\beta), \end{aligned}$$

which proves the second part of the lemma.

The lemma actually holds for  $0 < \beta \leq \frac{1}{2}$ , since trivially

$$\left| \int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha \right| \leq 2\beta |S(0)|^2 \ll \beta N^2,$$

which for  $0 < \beta \leq 1/N$  is  $\ll \beta^{-1/2} N^{1/2}$ .

**3. Proof of the Theorem.** Let  $T(\alpha) = \sum_{n=-K}^K t(n)e(-n\alpha)$  be any trigonometric polynomial with the properties that the  $t(n)$  are real numbers,  $t(-n) = t(n)$ , and  $T(\alpha) \geq 0$ . By equation (17) we have

$$\sum_{n=-K}^K t(n)R(N, N + n) = \int_{-\beta}^{\beta} S(\alpha)^2 T(\alpha) e(-N\alpha) d\alpha + \int_{\beta}^{1-\beta} S(\alpha)^2 T(\alpha) e(-N\alpha) d\alpha. \tag{21}$$

The second integral above may be estimated in terms of  $Z(N; n)$ ; we have

$$\begin{aligned} \left| \int_{\beta}^{1-\beta} S(\alpha)^2 T(\alpha) e(-N\alpha) d\alpha \right| &\leq \int_{\beta}^{1-\beta} |S(\alpha)|^2 T(\alpha) d\alpha \\ &= \int_0^1 |S(\alpha)|^2 T(\alpha) d\alpha - \int_{-\beta}^{\beta} |S(\alpha)|^2 T(\alpha) d\alpha \\ &= t(0)(N \log N + O(N)) + 2 \sum_{n=1}^K t(n)Z(N; n) \\ &\quad - \int_{-\beta}^{\beta} |S(\alpha)|^2 T(\alpha) d\alpha, \end{aligned} \tag{22}$$

where we have used the prime number theorem in the last line. In Lemma 1 take  $\beta = N^{-1/3}$ , and suppose  $n = N + o(N)$ . Then, since  $J(N, h) = o(h^2N)$  for  $h \geq N^{1/6+\epsilon}$  as mentioned in the introduction, we have  $J(N, 1/2\beta) = o(N/\beta^2)$ , and hence we have

$$\int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha \sim N, \quad \int_{-\beta}^{\beta} |S(\alpha)|^2 e(-n\alpha) d\alpha \sim N. \tag{23}$$

By (21), (22), and (23) we conclude that for  $K = o(N)$ ,

$$\begin{aligned} \left| \sum_{n=-K}^K t(n)R(N, N+n) - T(0)N \right| &\leq t(0)(N \log N + O(N)) + 2 \sum_{n=1}^K t(n)Z(N; n) \\ &\quad - T(0)N + o(T(0)N). \end{aligned} \tag{24}$$

The simplest choice for  $T(\alpha)$  is to take  $t(n) = K - |n|$ , which gives

$$T(\alpha) = \left( \frac{\sin \pi K \alpha}{\sin \pi \alpha} \right)^2,$$

and (24) becomes

$$\begin{aligned} \left| \sum_{n=-K}^K (K - |n|)R(N, N+n) - K^2N \right| &\leq KN \log N + 2 \sum_{n=1}^K (K - n)Z(N; n) \\ &\quad - K^2N + O(KN) + o(K^2N). \end{aligned}$$

The bound just obtained may be expressed in terms of  $J(N, K)$ . We have, for  $K \leq N/\log^3 N$ , and using the prime number theorem,

$$\begin{aligned} J(N, K) &= \int_1^N (\vartheta(x+K) - \vartheta(x))^2 dx - 2K \int_1^N (\vartheta(x+K) - \vartheta(x)) dx + K^2(N-1) \\ &= \int_1^N \left( \sum_{x < p \leq x+K} \log p \right)^2 dx - 2K^2 \sum_{p \leq N} \log p + O(K^3 \log N) + K^2N \\ &= K \sum_{p \leq N} \log^2 p + 2 \sum_{\substack{p, p' \leq N \\ 0 < p' - p \leq K}} \log p \log p' (K - (p' - p)) \\ &\quad - K^2N + O(K^3 \log^2 N) + o(K^2N) \\ &= K(N \log N + O(N)) + 2 \sum_{n=1}^K (K - n)Z(N; n) - K^2N + o(K^2N). \end{aligned} \tag{25}$$

We complete the proof of the Theorem by noting the trivial estimate  $R(N, N + n) = R(N + n) + O(n \log^2 N)$  allows us to replace  $R(N, N + n)$  by  $R(N + n)$  with an error of  $O(K^3 \log^2 N) = o(K^2 N)$  since  $K \leq N/\log^3 N$ .

**4. Comments on Linnik’s method.** Linnik proved assuming the Riemann Hypothesis that there is a Goldbach number in every interval  $[N, N + (\log N)^{3+\epsilon}]$ , for  $N$  sufficiently large. This result is weaker than the result we obtain where it suffices to have intervals of length  $C(\log N)^2$ , where  $C$  is a sufficiently large constant. The improvement results from two changes in Linnik’s original proof. First, Linnik does not work with  $S(\alpha)$ , but rather with

$$S(\vartheta, N) = \sum_{n=2}^{\infty} e^{-n/N} \Lambda(n) e(-\vartheta n),$$

where  $\Lambda(n) = \log p$  if  $n = p^m$ ,  $m \geq 1$ , and  $\Lambda(n) = 0$  otherwise. The reason for using  $S(\vartheta, N)$  is that it allows the use of an explicit formula where the sum over zeros of the Riemann zeta-function is absolutely convergent. This replaces the need to quote results on  $J(N, h)$ . The prime powers have no effect on the argument, but the fact that  $S(\vartheta, N)$  does not truncate at  $N$  but rather smoothly decreases causes the loss of a factor of  $\log N$  in Lemma 1.

The second difference in the argument is in the treatment of the second term in (17). Linnik’s argument is that for any continuous function  $V(\alpha)$  we have

$$\begin{aligned} \left| \int_{\beta}^{1-\beta} S(\alpha)^2 V(\alpha) e(-N\alpha) d\alpha \right| &\leq \max_{\beta \leq \alpha \leq 1-\beta} |V(\alpha)| \int_0^1 |S(\alpha)|^2 d\alpha \\ &= \max_{\beta \leq \alpha \leq 1-\beta} |V(\alpha)| \sum_{p \leq N} \log^2 p \\ &\ll \max_{\beta \leq \alpha \leq 1-\beta} |V(\alpha)| N \log N. \end{aligned}$$

The choice

$$V(\alpha) = \left( \sum_{k=1}^K e(-k\alpha) \right)^H = \sum_{k=1}^{HK} c(k) e(-k\alpha),$$

where  $H$  is a positive integer, satisfies the bound

$$\max_{\beta \leq \alpha \leq 1-\beta} |V(\alpha)| \ll \frac{1}{\beta^H}.$$

Assuming the Riemann Hypothesis, Selberg’s estimate (4) implies Lemma 1 gives an asymptotic relation provided  $\beta \log^2 N \rightarrow 0$ . Thus, applying (17) we obtain

$$\sum_{k=1}^{HK} c(k) R(N, N + k) \sim K^H N,$$

provided that  $\beta \log^2 N \rightarrow 0$  and  $N \log N / \beta^H = o(K^H N)$ . These conditions imply that  $K / \log^{2+1/H} N \rightarrow \infty$ , and since  $H$  can be taken arbitrarily large, we obtain Goldbach numbers in intervals of length  $(\log N)^{2+\epsilon}$ . While there are other possible choices of  $V(\alpha)$ , this argument always seems to lose something in the final result.

The success of Linnik's method depends on a special averaging property of the generating function. If we write (17) in the form

$$R(N, n) = \int_{-\beta}^{\beta} S(\alpha)^2 e(-n\alpha) d\alpha + \int_{\beta}^{1-\beta} S(\alpha)^2 e(-n\alpha) d\alpha = I_1 + I_2,$$

then we might expect that  $I_1$  is an approximation of  $R(N, n)$ . However, this turns out to be false. By Lemma 1 we have for appropriate ranges of  $\beta$  and  $n$  that  $I_1 \sim N$ . On the other hand, for odd,  $n$   $R(N, n)$  is either 0 or  $O(\log N)$ , while for even  $n$  we conjecture  $R(N, n) \sim \mathfrak{S}(n)N$  for  $n \sim N$ . By equation (28) in the next section  $\mathfrak{S}(n)$  is on average 1, which implies that on average for even  $n$  we have  $R(N, n) \sim 2N$ . Since  $I_1$  can not distinguish between even and odd numbers  $n$ , it takes on the average value of the two cases. We therefore get the correct expected value for  $R(N, n)$  from  $I_1$  when we average over both even and odd  $n$  in a sufficiently long interval.

**5. The corollaries.** To prove Corollary 1, assume  $J(N, K) = o(K^2N)$  in the Theorem. Then we have

$$\sum_{n=-K}^K (K - |n|)R(N + n) \sim K^2N \quad \text{as } N \rightarrow \infty.$$

Denote the sum above by  $S(K)$ . We have, for  $1 \leq h \leq K$ ,

$$\frac{S(K) - S(K - h)}{h} \leq \sum_{n=-K}^K R(N + n) \leq \frac{S(K + h) - S(K)}{h}$$

and hence

$$\left| \sum_{n=-K}^K R(N + n) - 2KN \right| < hN + o(K^2N/h).$$

On letting  $h = f(K)K$  where  $f(K) \rightarrow 0$  sufficiently slowly as  $k \rightarrow \infty$ , we have

$$\sum_{n=-K}^K R(N + n) \sim 2KN \quad \text{as } N \rightarrow \infty.$$

We have proved this subject to the bound  $K \leq N/\log^3 N$  from the Theorem. In this range we have  $R(N + n) \sim \log^2 N r(N + n)$ , which proves Corollary 1 after a relabeling of the variables. The restriction on the size of  $K$  may now be dropped by summing the result in (7).

To prove Corollary 2, we note that by Cauchy's inequality,

$$\left( \sum_{n=N}^{N+K} r(n) \right)^2 \left( \sum_{n=N}^{N+K} r(n)^2 \right)^{-1} \leq \sum_{\substack{N \leq n \leq N+K \\ r(n) \geq 1}} 1.$$

By the sieve bound  $r(n) \ll \mathfrak{S}(n)n/\log^2 n$  corresponding to (14) and by Corollary 1 we have that the left side above is

$$\gg K^2 \left( \sum_{n=N}^{N+K} \mathfrak{S}(n)^2 \right)^{-1},$$



if  $J(N, K) = o(K^2N)$ . Corollary 2 now follows from the estimate

$$\sum_{n=N}^{N+K} \mathfrak{S}(n)^2 \ll K + \log^2 N, \tag{26}$$

which can be proved by obvious modifications in the usual method for proving the result when  $K = N$  (see [9, p. 527]). Equation (26) is also an immediate consequence of the following result which may have some independent interest.

LEMMA 2. *We have, as  $N \rightarrow \infty$ ,*

$$\sum_{n \leq N} \mathfrak{S}(n)^2 = 2 \prod_{p \geq 3} \left(1 + \frac{1}{(p-1)^3}\right) N + O(\log^2 N). \tag{27}$$

The product above is the one that occurs in the Goldbach problem for representing an odd number as the sum of three primes. (This is not unexpected in view of the conjectures in (12).) The proof of Lemma 2 is nearly identical to the proof of the corresponding result

$$\sum_{n \leq N} \mathfrak{S}(n) = N + O(\log N) \tag{28}$$

from [13, Lemma 17.4].

*Proof of Lemma 2.* Let

$$f(j) = \begin{cases} \mu^2(j) \prod_{p|j} \frac{2p-3}{(p-2)^2}, & (j, 2) = 1; \\ 0, & (j, 2) = 2. \end{cases} \tag{29}$$

Then we have

$$\begin{aligned} \sum_{n \leq N} \mathfrak{S}(n)^2 &= 4C^2 \sum_{\substack{n \leq N \\ n \text{ even}}} \prod_{\substack{p|n \\ p \geq 3}} \left(\frac{p-1}{p-2}\right)^2 = 4C^2 \sum_{2m \leq N} \prod_{\substack{p|m \\ p \geq 3}} \left(1 + \frac{2p-3}{(p-2)^2}\right) \\ &= 4C^2 \sum_{m \leq N/2} \sum_{j|m} f(j) = 4C^2 \sum_{j \leq N/2} f(j) \left[\frac{N}{2j}\right] \\ &= 2C^2 N \sum_{j=1}^{\infty} \frac{f(j)}{j} + O\left(N \sum_{j \geq N} \frac{f(j)}{j}\right) + O\left(\sum_{j \leq N} f(j)\right). \end{aligned}$$

First, we have

$$\begin{aligned} 2C^2 \sum_{j=1}^{\infty} \frac{f(j)}{j} &= 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right)^2 \left(1 + \frac{2p-3}{p(p-2)^2}\right) \\ &= 2 \prod_{p \geq 3} \left(1 + \frac{1}{(p-1)^3}\right). \end{aligned}$$

To estimate the error terms, note

$$f(j) \ll \prod_{\substack{p|j \\ p \geq 3}} \frac{2p}{(p-2)^2} = \prod_{\substack{p|j \\ p \geq 3}} \left(\frac{2}{p}\right) \prod_{\substack{p|j \\ p \geq 3}} \frac{1}{(1-2/p)^2} \ll \prod_{p|j} \left(\frac{2}{p}\right) \prod_{p|j} \left(1 + \frac{4}{p}\right).$$

For  $j$  squarefree we have

$$\prod_{p|j} \frac{2}{p} = \frac{2^{\omega(j)}}{j} = \frac{d(j)}{j},$$

$$\prod_{p|j} \left(1 + \frac{4}{p}\right) = \sum_{h|j} \frac{4^{\omega(h)}}{h} = \sum_{h|j} \frac{d(h)^2}{h},$$

where  $\omega(n)$  is the number of distinct prime factors of  $n$ , and  $d(n)$  is the number of divisors of  $n$ . For  $j$  not squarefree  $f(j) = 0$ , and hence in general

$$f(j) \ll \frac{d(j)}{j} \sum_{h|j} \frac{d(h)^2}{h}. \tag{30}$$

We recall [9, Theorem 5.3] that, for  $q$  a non-negative integer,

$$\sum_{n \leq N} d(n)^q \ll N(\log N)^{2q-1}, \quad \sum_{n \leq N} \frac{d(n)^q}{n} \ll (\log N)^{2q}.$$

Hence by (30)

$$\begin{aligned} \sum_{j \leq N} f(j) &\ll \sum_{j \leq N} \frac{d(j)}{j} \sum_{h|j} \frac{d(h)^2}{h} \ll \sum_{h \leq N} \frac{d(h)^2}{h} \sum_{k \leq N/h} \frac{d(hk)}{hk} \\ &\ll \sum_{h \leq N} \frac{d(h)^3}{h^2} \sum_{k \leq N} \frac{d(k)}{k} \ll \log^2 N, \end{aligned}$$

and

$$\sum_{j \geq N} \frac{f(j)}{j} \ll \sum_{k=0}^{\infty} \frac{1}{2^k N} \sum_{2^k N \leq j < 2^{k+1} N} f(j) \ll \frac{\log^2 N}{N},$$

which proves Lemma 2.

A. Ivić has pointed out to me the following argument which improves on Corollary 2. By Hölder’s inequality we have, for any number  $\nu > 1$ ,

$$\sum_{n=N}^{N+K} r(n) \leq \left( \sum_{n=N}^{N+K} r(n)^\nu \right)^{1/\nu} \left( \sum_{\substack{N \leq n \leq N+K \\ r(n) \geq 1}} 1 \right)^{1-1/\nu}.$$

As before we obtain, if  $J(N, K) = o(K^2 N)$ ,

$$\sum_{\substack{N \leq n \leq N+K \\ r(n) \geq 1}} 1 \gg K^{\nu/(\nu-1)} \left( \sum_{n=N}^{N+K} \mathfrak{S}(n)^\nu \right)^{-1/(\nu-1)}.$$

Lemma 2 may be replaced by

$$\sum_{n \leq N} \mathfrak{S}(n)^\nu = C(\nu)N + O_\nu(\log^\nu N), \tag{31}$$

for  $\nu \geq 1$ . The proof of equation (31) is similar to the proof of Lemma 2; the error terms may be estimated by using Theorem 14.9 of [10]. On differencing (31) we may replace (26) by

$$\sum_{n=N}^{N+K} \mathfrak{S}(n)^\nu \ll_\nu K + \log^\nu N, \quad \nu \geq 1,$$

and Corollary 2 follows for  $K \geq C \log^\nu N$ ,  $\nu > 1$ .

To prove Corollary 3, we use (14) in (25) and obtain

$$J(N, K) \leq KN \log N + 2(A + \epsilon)N \sum_{n=1}^K (K - n)\mathfrak{S}(n) - K^2N + o(K^2N) + O(NK).$$

By [4, Lemma 3] (with a change of notation),

$$\sum_{n=1}^K (K - n)\mathfrak{S}(n) = \frac{1}{2}K^2 - \frac{1}{2}K \log K + O(K), \tag{32}$$

which gives

$$J(N, K) \leq (A - 1 + \epsilon)K^2N + KN \log N + O(KN \log K) + o(K^2N).$$

By the Theorem we conclude that  $R(n)$  is positive for some values of  $n$  in  $[N - K, N + K]$  if  $(A - 1 + \epsilon)K^2N + KN \log N < (1 - \epsilon)K^2N$  which holds if  $K > \frac{1}{2 - A - 2\epsilon} \log N$ .

Corollary 3 follows on replacing  $K$  by  $K/2$ .

**6. A Conjecture on  $J(N, h)$ .** It seems worthwhile to mention that (32) may be replaced by a more precise result, namely

$$\sum_{n=1}^K (K - n)\mathfrak{S}(n) = \frac{1}{2}K^2 - \frac{1}{2}K \log K - \frac{1}{2}(-1 + \gamma + \log 2\pi)K + O(K^{1/2+\epsilon}), \tag{33}$$

where  $\gamma$  is Euler’s constant. The proof uses standard methods, and will not be given since (33) has no application to the problem considered in this paper. However, the relation in (33) does have an interesting heuristic connection with  $J(N, h)$ . Suppose that the conjecture stated in (12) is strengthened to

$$Z(N; n) = \mathfrak{S}(n)N + O(N^{1/2+\epsilon}), \quad 1 \leq n \leq N. \tag{34}$$

By using the prime number theorem with the error term implied by the Riemann Hypothesis, equation (25) may be replaced by

$$J(N, K) = KN \log N - KN + 2 \sum_{n=1}^K (K - n)Z(N; n) - K^2N + O(K^2N^{1/2+\epsilon}) + O(K^3 \log N).$$

On using (34) and then (33) we conclude

$$J(N, K) = KN \log(N/K) - (\gamma + \log 2\pi)KN + O(K^2N^{1/2+\epsilon}) + O(K^{1/2+\epsilon}N) + O(K^3 \log N).$$

Thus, subject to (34) and the Riemann Hypothesis, we have

$$J(N, K) = KN \log(N/K) - (\gamma + \log 2\pi)KN + o(KN) + O(N), \tag{35}$$

uniformly for  $1 \leq K \leq N^{1/2-\epsilon}$ . It is reasonable to conjecture that this relation continues to hold for larger values of  $K$ , and also for non-integer values of  $K$ . We thus make the following conjecture.

CONJECTURE. *We have*

$$J(N, h) = hN \log(N/h) - (\gamma + \log 2\pi)hN + o(hN) + O(N) \tag{36}$$

*uniformly for  $1 \leq h \leq N^{1-\epsilon}$ .*

The conjecture implies that there always exist primes between  $n^2$  and  $(n+1)^2$ , for all sufficiently large  $n$ . Furthermore, using the analysis in [5], one finds heuristically that (36) implies

$$F(X, T) = \frac{T}{2\pi} \log T - (1 + \log 2\pi) \frac{T}{2\pi} + o(T), \quad (37)$$

for  $T \leq X \leq T^M$ , where  $M$  is any fixed number  $\geq 1$ . Here  $F(X, T)$  is the sum over pairs of zeros of the Riemann zeta-function  $\zeta(s)$  introduced by Montgomery, see [5]. Let  $N(T)$  denote the number of zeros of  $\zeta(s)$ ,  $s = \sigma + it$ , in the rectangle  $0 < \sigma < 1$ ,  $0 < t < T$ . The first two terms in (37) are the same as the first two terms in the Riemann-von Mangoldt formula

$$N(T) = \frac{T}{2\pi} \log T - (1 + \log 2\pi) \frac{T}{2\pi} + O(\log T).$$

If we denote the number of simple zeros of  $\zeta(s)$  in the above rectangle by  $N_s(T)$ , one finds heuristically that (37) implies

$$N_s(T) = N(T) + o(T).$$

Thus, the conjecture is consistent with the expectation that all the zeros of the Riemann zeta-function are simple.

#### REFERENCES

1. J. R. Chen, On the Goldbach's problem and the sieve methods, *Sci. Sinica* **21** (1978), 701–739.
2. E. Fouvry and F. Grupp, On the switching principle in sieve theory, *J. Reine Angew. Math.* **370** (1986), 101–126.
3. P. X. Gallagher, A large sieve density estimate near  $\sigma = 1$ , *Invent. Math.* **11** (1970), 329–339.
4. D. A. Goldston, The second moment for prime numbers, *Quart. J. Math. Oxford* (2) **35** (1984), 153–163.
5. D. A. Goldston and H. L. Montgomery, Pair correlation of zeros and primes in short intervals, *Analytic Number Theory and Diophantine Problems*, (Birkhäuser 1987), 183–203.
6. H. Halberstam and H.-E. Richert, *Sieve Methods* (Academic Press, 1974).
7. G. H. Hardy and J. E. Littlewood, Some Problems of 'Partitio Numerorum': III, On the expression of a number as a sum of primes, *Acta. Math.* **44** (1923), 1–70.
8. D. R. Heath-Brown, Primes in 'almost all' short intervals, *J. London Math. Soc.* (2) **26** (1982), 385–396.
9. Hua Loo Keng, *Introduction to Number Theory* (Springer-Verlag, 1982).
10. Aleksandar Ivić, *The Riemann Zeta-Function* (John Wiley and Sons, 1985).
11. I. Kátai, A remark on a paper of Ju. V. Linnik, *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **17** (1967), 99–100.
12. Yu. V. Linnik, Some conditional theorems concerning the binary Goldbach problem, *Izv. Akad. Nauk SSSR* **16** (1952), 503–520.
13. H. L. Montgomery, *Topics in Multiplicative Number Theory*, Lecture Notes in Mathematics 227 (Springer-Verlag, 1971).
14. H. L. Montgomery and R. C. Vaughan, The exceptional set in Goldbach's problem, *Acta Arith.* **27** (1975), 353–370.

15. K. Ramachandra, On the number of Goldbach numbers in small intervals, *J. Indian Math. Soc.* **37** (1973), 157–170.
16. K. Ramachandra, Two remarks in prime number theory, *Bull. Soc. Math. France* **105** (1977), 433–437.
17. A. Selberg, On the normal density of primes in small intervals, and the difference between consecutive primes, *Arch. Math. Naturvid.* **47** (1943), no. 6, 87–105.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE  
SAN JOSE STATE UNIVERSITY  
SAN JOSE, CA 95192  
USA