PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 95, Number 4, December 1985

ON A CONJECTURE OF BALOG

ADOLF HILDEBRAND

ABSTRACT. A conjecture of A. Balog is proved which gives a sufficient condition on a set A of positive integers such that $A \cap (A + 1)$ is infinite. A consequence of this result is that, for every $\varepsilon > 0$, there are infinitely many integers n such that both n and $n + 1$ have a prime factor $> n^{1-\varepsilon}$.

1. Introduction. Some of the most difficult and seemingly unattackable problems in number theory deal with simultaneous properties of integers n and their translates $n + t$, where $t \in \mathbb{N}$ is fixed. The twin prime conjecture, for example, asserts that n and $n + 2$ are prime infinitely often.

Another problem of this type, posed by Erdös several times (see e.g. [3]), is to show that, for every fixed $\varepsilon > 0$, there are infinitely many integers n such that both n and $n + 1$ have a prime factor $> n^{1-\varepsilon}$. In other words, putting

$$Q_\alpha = \{n \in \mathbb{N}: P(n) > n^\alpha\},$$

where $P(n)$ denotes the largest prime factor of n, the conjecture asserts that $Q_\alpha \cap (Q_\alpha + 1)$ is infinite for every $\alpha < 1$.

At the Oberwolfach meeting on analytic number theory in 1982, A. Balog proposed a general conjecture, which gives a sufficient condition on a set $A \subset \mathbb{N}$, such that $A \cap (A + 1)$ contains infinitely many elements. To this end, he introduced the concept of "k-stability". A set $A \subset \mathbb{N}$ is called k-stable if

$$kA \mathrel{\dot{\subset}} A, \qquad k^{-1}(A \cap k\mathbb{N}) \mathrel{\dot{\subset}} A,$$

where λA denotes the set $\{\lambda a: a \in A\}$, and $A \mathrel{\dot{\subset}} B$ means that A is contained in B up to a set of density zero, i.e., $\mathrm{d}(A \setminus B) = 0$. Here and in the sequel, $\mathrm{d}(\cdot)$ denotes the asymptotic density, defined by

$$\mathrm{d}(A) = \lim_{x \to \infty} \frac{1}{x} \sum_{\substack{n \leq x \\ n \in A}} 1$$

(provided this limit exists), and the lower and upper densities $\underline{\mathrm{d}}(\cdot)$ and $\overline{\mathrm{d}}(\cdot)$ are defined analogously by taking the limit inferior and the limit superior, respectively.

Balog [1] showed by an elementary argument that $A \cap (A + 1)$ is infinite whenever A is 2-stable and $\underline{\mathrm{d}}(A) > 1/3$, and he made the following

CONJECTURE (BALOG [1]). If $A \subset \mathbb{N}$ is p-stable for every prime p and has positive density, then $A \cap (A + 1)$ is infinite.

Received by the editors December 14, 1984 and, in revised form, February 4, 1985.
1980 *Mathematics Subject Classification*. Primary 10A50, 10L99; Secondary 10H15.

The sets Q_α, $\alpha < 1$, introduced above, have positive density (see e.g. [**2**]), and it is easy to see that they are k-stable for every $k \in \mathbb{N}$. Thus Balog's conjecture implies the above-mentioned conjecture that $P(n) > n^{1-\varepsilon}$ and $P(n+1) > (n+1)^{1-\varepsilon}$ holds infinitely often for every fixed $\varepsilon > 0$.

The purpose of this paper is to prove Balog's conjecture in a more general form.

2. Results. Given a set $A \subset \mathbb{N}$, we define, for $N \in \mathbb{N}$,

$$A_N = \bigcup_{n,d=1}^{N} \frac{n}{d}(A \cap d\mathbb{N}).$$

These sets form an ascending chain starting with $A_1 = A$.

THEOREM. *If $\underline{\mathrm{d}}(A) > 0$, then $\underline{\mathrm{d}}(A_N \cap (A_N + 1)) > 0$ for all sufficiently large N. More precisely, for every $\varepsilon > 0$ there exist $N(\varepsilon) \in \mathbb{N}$ and $\delta(\varepsilon) > 0$ such that $\underline{\mathrm{d}}(A) \geqslant \varepsilon$ implies*

$$\underline{\mathrm{d}}(A_N \cap (A_N + 1)) \geqslant \delta(\varepsilon) \qquad (N \geqslant N(\varepsilon)).$$

If A is p-stable for every prime $p \leqslant N$, then A is k-stable for all $k \leqslant N$, so that

$$A_N = \bigcup_{n,d=1}^{N} \frac{n}{d}(A \cap d\mathbb{N}) \subset A \subset A_N,$$

and, therefore, $\underline{\mathrm{d}}(A_N \cap (A_N + 1)) = \underline{\mathrm{d}}(A \cap (A + 1))$. Thus the theorem implies Balog's conjecture in the following form.

COROLLARY 1. *If $A \subset \mathbb{N}$ satisfies $\underline{\mathrm{d}}(A) \geqslant \varepsilon$ and is p-stable for every prime $p \leqslant N(\varepsilon)$, then $\underline{\mathrm{d}}(A \cap (A + 1)) \geqslant \delta(\varepsilon)$ holds, where $N(\varepsilon)$ and $\delta(\varepsilon)$ are as in the Theorem.*

Applying this result to the sets $Q_\alpha \setminus Q_\beta$, $0 \leqslant \alpha < \beta \leqslant 1$, we obtain the conjecture mentioned in the introduction in the following slightly more general form.

COROLLARY 2. *Let $0 \leqslant \alpha < \beta \leqslant 1$. Then the set of integers n for which $n^\alpha < P(n) \leqslant n^\beta$, $(n+1)^\alpha < P(n+1) \leqslant (n+1)^\beta$ holds has positive lower density.*

3. Lemmas.

LEMMA 1. *For every $k \geqslant 2$ there exist positive integers $n_1 < \cdots < n_k$ satisfying*

$$n_j - n_i = (n_i, n_j) \qquad (1 \leqslant i < j \leqslant k). \tag{1}$$

PROOF.[1] We define an auxiliary sequence $(N_k)_{k \geqslant 1}$ recursively by

$$N_1 = 1, \quad N_{k+1} = 2\prod_{i=1}^{k}\left(\sum_{j=i}^{k} N_j\right) \qquad (k \geqslant 1).$$

By construction,

$$\sum_{h=i}^{j-1} N_h | N_j | N_{j+1} | \cdots | N_k \qquad (1 \leqslant i < j \leqslant k).$$

[1] Heath-Brown [**4**] proved a stronger form of the lemma, where the n_i were required to satisfy an additional condition besides (1). Since this additional condition is of no relevance here and complicates the proof considerably, we preferred to give a short proof of the lemma in the form stated.

Thus, if for given $k \geqslant 2$ we put

$$n_k = N_k, \quad n_i = N_k - \sum_{j=i}^{k-1} N_j \qquad (1 \leqslant i \leqslant k-1),$$

we have $n_k > \cdots > n_1 \geqslant N_k/2 \geqslant 1$, and

$$n_j - n_i = \sum_{h=i}^{j-1} N_h | N_k - \sum_{h=j}^{k-1} N_h = n_j \qquad (1 \leqslant i < j \leqslant k),$$

which is equivalent to (1).

LEMMA 2. *Let r be a positive integer, and for $D \geqslant 1$ let $\mathcal{D} = \mathcal{D}(D, r)$ be the set of positive integers $d \leqslant D$ of the form*

$$d = d_1 p, \quad (p, r) = 1, \quad q | d_1 \Rightarrow q | r, \tag{2}$$

where p and q denote primes. Then we have

$$\frac{1}{r} \sum_{d \in \mathcal{D}} \frac{1}{d} = \frac{\log\log(D+2)}{\varphi(r)} + O(1) \tag{3}$$

and

$$\mathrm{d}\Big(\mathbb{N} \setminus \bigcup_{d \in \mathcal{D}} d(r\mathbb{N} - 1)\Big) \ll \frac{\varphi(r)}{\log\log(D+2)}, \tag{4}$$

where φ is the Euler function and the implied constants are absolute.

PROOF. Letting d_1 be an integer all of whose prime factors divide r, we have

$$\begin{aligned}\sum_{d \in \mathcal{D}} \frac{1}{d} &\leqslant \sum_{d_1 \leqslant D} \frac{1}{d_1} \sum_{p \leqslant D} \frac{1}{p} \leqslant \prod_{p | r}\left(1 - \frac{1}{p}\right)^{-1} \sum_{p \leqslant D} \frac{1}{p} \\ &= \frac{r}{\varphi(r)}(\log\log(D+2) + O(1))\end{aligned}$$

and

$$\begin{aligned}\sum_{d \in \mathcal{D}} \frac{1}{d} &\geqslant \sum_{d_1 \leqslant \sqrt{D}} \frac{1}{d_1} \sum_{\substack{p \leqslant \sqrt{D} \\ p \nmid r}} \frac{1}{p} \\ &\geqslant \left(\frac{r}{\varphi(r)} - \sum_{d_1 > \sqrt{D}} \frac{1}{d_1}\right) \log\log(D+2) + O(r).\end{aligned}$$

This yields (3), since

$$\begin{aligned}\sum_{d_1 > \sqrt{D}} \frac{1}{d_1} &\leqslant D^{-1/4} \sum d_1^{-1/2} = D^{-1/4} \prod_{p | r} (1 - p^{-1/2})^{-1} \\ &\ll D^{-1/4} \exp\Big(2 \sum_{p | r} p^{-1/2}\Big) \ll r D^{-1/4}.\end{aligned}$$

For the proof of (4) we may suppose

$$\log\log D \geqslant C\varphi(r), \tag{5}$$

where C is an arbitrary, but fixed, positive constant. Set $S = \bigcup_{d \in \mathcal{D}} d(r\mathbb{N} - 1)$ and define

$$f(n) = \sum_{\substack{d \mid n,\, d \in \mathcal{D} \\ n/d \equiv -1 \bmod r}} 1.$$

Thus $f(n) \geqslant 0$ for all $n \in \mathbb{N}$, and $f(n) > 0$ if and only if $n \in S$. We use a variance argument to obtain the desired upper bound for $\mathrm{d}(\mathbb{N} \setminus S)$.

Putting

$$(6) \qquad M = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leqslant x} f(n) = \sum_{d \in \mathcal{D}} \frac{1}{d} \lim_{x \to \infty} \frac{d}{x} \sum_{\substack{n \leqslant x/d \\ n \equiv -1 \bmod r}} 1 = \frac{1}{r} \sum_{d \in \mathcal{D}} \frac{1}{d},$$

we have

$$\begin{aligned}
\mathrm{d}(\mathbb{N} \setminus S) &= \lim_{x \to \infty} \frac{1}{x} \sum_{\substack{n \leqslant x \\ f(n) = 0}} 1 \leqslant \lim_{x \to \infty} \frac{1}{x} \sum_{\substack{n \leqslant x \\ |f(n) - M| \geqslant M/2}} 1 \\
&\leqslant \frac{4}{M^2} \lim_{x \to \infty} \frac{1}{x} \sum_{n \leqslant x} (f(n) - M)^2 = \frac{4}{M^2} \left\{ \lim_{x \to \infty} \frac{1}{x} \sum_{n \leqslant x} f(n)^2 - M^2 \right\} \\
&= \frac{4}{M^2} \{ M_2 - M^2 \}, \quad \text{say.}
\end{aligned}$$

In view of (3), (6), and (5) (with a sufficiently large constant C), the asserted upper bound (4) follows if we can show

$$(7) \qquad M_2 = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leqslant x} f(n)^2 \leqslant \left(\frac{\log \log D}{\varphi(r)} \right)^2 + O\left(\frac{\log \log D}{\varphi(r)} \right)$$

with an absolute O-constant.

Expanding $f(n)^2$, we get

$$M_2 = \sum_{d, d' \in \mathcal{D}} \frac{1}{[d, d']} \lim_{x \to \infty} \frac{[d, d']}{x} \sum_{\substack{n \leqslant x/[d, d'] \\ (*)}} 1,$$

where $(*)$ denotes the condition

$$(*) \qquad \frac{nd}{(d, d')} \equiv \frac{nd'}{(d, d')} \equiv -1 \bmod r.$$

Let $d = d_1 p$ and $d' = d_1' p'$ be the (unique) decompositions of the form (2) for d and d'. Then $(*)$ has a solution in n if and only if $d_1 = d_1'$, $p \equiv p' \bmod r$, and in this case the limit in the last expression equals $1/r$. Thus we get

$$\begin{aligned}
M_2 &\leqslant \frac{1}{r} \sum_{d_1 \leqslant D} \frac{1}{d_1} \sum_{\substack{p, p' \leqslant D \\ p \equiv p' \bmod r \\ p \nmid r}} \frac{1}{[p, p']} \\
&\leqslant \frac{1}{\varphi(r)} \sum_{\substack{p \leqslant D \\ p \nmid r}} \frac{1}{p} \left(\sum_{\substack{p' \leqslant D \\ p' \equiv p \bmod r}} \frac{1}{p'} + O(1) \right).
\end{aligned}$$

The innermost sum equals

$$\sum_{\substack{e^r \leqslant p' \leqslant D \\ p' \equiv p \bmod r}} \frac{1}{p'} + O\left(\sum_{\substack{n \leqslant e^r \\ n \equiv p \bmod r}} \frac{1}{n} \right)$$

$$= \int_{e^r}^{D} \pi(x, r, p) \frac{dx}{x^2} + O(1) = \frac{\log\log D}{\varphi(r)} + O(1),$$

where the last step follows from the Siegel-Walfisz theorem. We therefore obtain

$$M_2 \leqslant \frac{1}{\varphi(r)} \left(\sum_{p \leqslant D} \frac{1}{p} \right) \left(\frac{\log\log D}{\varphi(r)} + O(1) \right)$$

$$\leqslant \left(\frac{\log\log D}{\varphi(r)} \right)^2 + O\left(\frac{\log\log D}{\varphi(r)} \right),$$

i.e., estimate (7). This completes the proof of Lemma 2.

4. Proof of the Theorem. For $x > 0$ let $\mathrm{d}_x(\cdot)$ be defined by

$$\mathrm{d}_x(M) = \frac{1}{x} \sum_{\substack{n \leqslant x \\ n \in M}} 1 \qquad (M \subset \mathbb{N}),$$

so that

$$\mathrm{d}(M) = \liminf_{x \to \infty} \mathrm{d}_x(M), \qquad \bar{\mathrm{d}}(M) = \limsup_{x \to \infty} \mathrm{d}_x(M).$$

If $\lambda \geqslant 1$, then obviously

$$\mathrm{d}_x(\lambda M) = (1/\lambda) \mathrm{d}_{x/\lambda}(M) \leqslant \mathrm{d}_x(M),$$

and for every fixed $t \in \mathbb{N}$ we have

$$\mathrm{d}_x(M + t) = \mathrm{d}_x(M) + o(1) \quad \text{as } x \to \infty.$$

Given a set $A \subset \mathbb{N}$ and positive integers $n_1 < \cdots < n_k$ satisfying (1), we define the sets

$$B_{i,d} = n_i(A + d) \cap nd\mathbb{N} \qquad (1 \leqslant i \leqslant k, d \in \mathbb{N}),$$

where $n = \prod_{i=1}^k n_i^2$. By the inclusion-exclusion principle we have, for $x > 0$ and every $d \in \mathbb{N}$,

$$\mathrm{d}_x\left(\bigcup_{i=1}^k B_{i,d} \right) \geqslant \sum_{i=1}^k \mathrm{d}_x(B_{i,d}) - \sum_{1 \leqslant i < j \leqslant k} \mathrm{d}_x(B_{i,d} \cap B_{j,d}). \tag{8}$$

We shall estimate from above the second term on the right in terms of

$$\mathrm{d}_x(A_N \cap (A_N + 1)),$$

where $N \geqslant \max(d, n)$, and bound the first term, averaged over a suitable range for d, from below in terms of $\mathrm{d}_{x/n_k}(A)$. This will lead to the desired relation between the densities of A and $A_N \cap (A_N + 1)$.

Using the stated properties of the function d_x, we obtain

$$\mathrm{d}_x(B_{i,d}) = \mathrm{d}_x\left(n_i\left((A + d) \cap d\frac{n}{n_i}\mathbb{N} \right) \right) = \frac{1}{n_i} \mathrm{d}_{x/n_i}\left((A + d) \cap d\frac{n}{n_i}\mathbb{N} \right)$$

$$\geqslant \frac{1}{n_k} \mathrm{d}_{x/n_k}(A \cap dT_i) + o(1),$$

where $T_i = (n/n_i)\mathbb{N} - 1$. Moreover, for $1 \leqslant i < j \leqslant k$, we get

$$\begin{aligned} \mathrm{d}_x(B_{i,d} \cap B_{j,d}) &\leqslant \mathrm{d}_x(n_i(A+d) \cap n_j(A+d) \cap dn_i n_j \mathbb{N}) \\ &\leqslant \mathrm{d}_x\left(\frac{n_i}{(n_i,n_j)} \frac{A \cap d\mathbb{N}}{d} \cap \left(\frac{n_j}{(n_i,n_j)} \frac{A \cap d\mathbb{N}}{d} + \frac{n_j - n_i}{(n_i,n_j)}\right)\right) + o(1) \\ &\leqslant \mathrm{d}_x(A_N \cap (A_N + 1)) + o(1), \end{aligned}$$

provided $N \geqslant \max(n, d)$, where the last step follows from (1) and the definition of A_N. Substituting these estimates together with the trivial bound

$$\mathrm{d}_x\left(\bigcup_{i=1}^{k} B_{i,d}\right) \leqslant \mathrm{d}_x(dn\mathbb{N}) \leqslant \frac{1}{dn}$$

into (8) yields

$$\frac{1}{dn} \geqslant \frac{1}{n_k} \sum_{i=1}^{k} \mathrm{d}_{x/n_k}(A \cap dT_i) - k^2 \mathrm{d}_x(A_N \cap (A_N + 1)) + o(1)$$

for every fixed $d \in \mathbb{N}$ and $N \geqslant \max(n, d)$.

We now fix $D \geqslant 1$ and let $\mathscr{D} = \mathscr{D}(D, n/n_i)$ be defined as in Lemma 2, with $r = n/n_i$. Since $\mathscr{D}(D, r)$ depends only on the set of prime factors of r, and the numbers $n/n_i = \prod_{j=1}^{k} n_j^2/n_i$, $1 \leqslant i \leqslant k$, have the same set of prime factors, this definition does not depend on the choice of the index i. Summing the last inequality over $d \in \mathscr{D}$, we obtain, for $N \geqslant \max(n, D)$,

$$\begin{aligned} \frac{1}{n} \sum_{d \in \mathscr{D}} \frac{1}{d} &\geqslant \frac{1}{n_k} \sum_{i=1}^{k} \sum_{d \in \mathscr{D}} \mathrm{d}_{x/n_k}(A \cap dT_i) - Dk^2 \mathrm{d}_x(A_N \cap (A_N + 1)) + o(1) \\ &\geqslant \frac{1}{n_k} \sum_{i=1}^{k} \mathrm{d}_{x/n_k}(A \cap S_i) - Dk^2 \mathrm{d}_x(A_N \cap (A_N + 1)) + o(1), \end{aligned}$$

where

$$S_i = \bigcup_{d \in \mathscr{D}} dT_i = \bigcup_{d \in \mathscr{D}} d\left(\frac{n}{n_i}\mathbb{N} - 1\right).$$

Letting $x \to \infty$, we deduce

$$\begin{aligned} (9) \qquad Dk^2 \underline{\mathrm{d}}(A_N \cap (A_N + 1)) &\geqslant \frac{1}{n_k} \sum_{i=1}^{k} \underline{\mathrm{d}}(A \cap S_i) - \frac{1}{n} \sum_{d \in \mathscr{D}} \frac{1}{d} \\ &\geqslant \frac{k}{n_k} \underline{\mathrm{d}}(A) - \frac{1}{n_k} \sum_{i=1}^{k} \mathrm{d}(\mathbb{N} \setminus S_i) - \frac{1}{n} \sum_{d \in \mathscr{D}} \frac{1}{d}. \end{aligned}$$

By Lemma 2 we have

$$\frac{n_k}{n} \sum_{d \in \mathscr{D}} \frac{1}{d} \ll \frac{\log\log(D+2)}{\varphi(n/n_k)} + 1$$

and

$$\mathrm{d}(\mathbb{N} \setminus S_i) \ll \frac{\varphi(n/n_i)}{\log\log(D+2)}.$$

Since

$$\frac{\varphi(n/n_i)}{n/n_i} = \prod_{p \mid n/n_i} \left(1 - \frac{1}{p}\right) = \prod_{p \mid n_1 \cdots n_k} \left(1 - \frac{1}{p}\right)$$

is independent of the choice of i, and since, in view of (1),

$$n_1 \leqslant n_i \leqslant n_k \leqslant 2n_1,$$

the last estimate remains valid with $\varphi(n/n_k)$ in place of $\varphi(n/n_i)$. Thus, defining $D = D(k)$ by

$$\frac{\log\log(D+2)}{\varphi(n/n_k)} = \sqrt{k},$$

we obtain, from (9),

$$Dk^2 \underline{\mathrm{d}}(A_N \cap (A_N + 1)) \geqslant \frac{k}{n_k}\left(\underline{\mathrm{d}}(A) + O\left(\frac{1}{\sqrt{k}}\right)\right)$$

with an absolute O-constant. If now $\underline{\mathrm{d}}(A) \geqslant \varepsilon$ (> 0), then by choosing $k = k(\varepsilon)$ sufficiently large (which is possible by Lemma 1), the O-term becomes $\leqslant \varepsilon/2$, and we get

$$\underline{\mathrm{d}}(A_N \cap (A_N + 1)) \geqslant \delta(\varepsilon) \qquad (N \geqslant N(\varepsilon)),$$

with

$$\delta(\varepsilon) = \frac{\varepsilon}{2D(k)kn_k}, \qquad N(\varepsilon) = \max(n, D(k)),$$

as asserted in the Theorem.

By a minor modification of the proof, one can show that the theorem remains valid, when $\underline{\mathrm{d}}$ is replaced by the upper density $\overline{\mathrm{d}}$.

ACKNOWLEDGEMENT. This work was done while the author was visiting the University of Illinois at Urbana-Champaign. I would like to thank the Department of Mathematics at Urbana for its hospitality and the Deutsche Forschungsgemeinschaft for financial support. Also, I am grateful to A. Balog for calling my attention to reference **[4]**.

REFERENCES

1. A. Balog, Problem in Tagungsbericht **41** (1982), Math. Forschungsinstitut Oberwolfach, p. 29.

2. N. G. de Bruijn, *On the number of positive integers $\leqslant x$ and free of prime factors $> y$*, Nederl. Akad. Wetensch. Proc. Ser. A **54** (1951), 50–60.

3. P. Erdös, *Problems and results on number theoretic properties of consecutive integers and related questions*, Proc. Fifth Manitoba Conf. on Num. Math. (Univ. Manitoba, Winnipeg, Man., 1975). Congressus Numeratium, no. XVI, Utilitas Math., Winnipeg, Man., 1976, pp. 25–44.

4. D. R. Heath-Brown, *The divisor function at consecutive integers*, Mathematika **31** (1984), 141–149.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801

Current address: School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540