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 \( A \subseteq \mathbb{N} \), such that \( A \cap (A + 1) \) contains infinitely many elements. To this end, he introduced the concept of “\( k \)-stability”. A set \( A \subseteq \mathbb{N} \) is called \( k \)-stable if
\[ kA \subseteq A, \quad k^{-1}(A \cap k\mathbb{N}) \subseteq A, \]
where \( \lambda A \) denotes the set \( \{ \lambda a : a \in A \} \), and \( A \subseteq B \) means that \( A \) is contained in \( B \) up to a set of density zero, i.e., \( d(A \setminus B) = 0 \). Here and in the sequel, \( d(\cdot) \) denotes the asymptotic density, defined by
\[ d(A) = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x, n \in A} 1 \]
(provided this limit exists), and the lower and upper densities \( d(\cdot) \) and \( \bar{d}(\cdot) \) 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 \( d(A) > 1/3 \), and he made the following

Conjecture (Balog [1]). If \( A \subseteq \mathbb{N} \) 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.

©1985 American Mathematical Society
0002-9939/85 $1.00 + $.25 per page

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
The sets $Q_{\alpha}, \alpha < 1$, 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 $P(n) > n^{1-\epsilon}$ and $P(n + 1) > (n + 1)^{1-\epsilon}$ holds infinitely often for every fixed $\epsilon > 0$.

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 $d(A) > 0$, then $d(A_N \cap (A_N + 1)) > 0$ for all sufficiently large $N$.

More precisely, for every $\epsilon > 0$ there exist $N(\epsilon) \in \mathbb{N}$ and $\delta(\epsilon) > 0$ such that $d(A) \geq \epsilon$ implies

$$d(A_N \cap (A_N + 1)) \geq \delta(\epsilon) \quad (N \geq N(\epsilon)).$$

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

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

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

**Corollary 1.** If $A \subset \mathbb{N}$ satisfies $d(A) \geq \epsilon$ and is $p$-stable for every prime $p \leq N(\epsilon)$, then $d(A \cap (A + 1)) \geq \delta(\epsilon)$ holds, where $N(\epsilon)$ and $\delta(\epsilon)$ are as in the Theorem.

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

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

3. Lemmas.

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

$$n_j - n_i = (n_i, n_j) \quad (1 \leq i < j \leq k).$$

**Proof.** We define an auxiliary sequence $(N_k)_{k \geq 1}$ recursively by

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

By construction,

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

---

1Heath-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 \( k \geq 2 \) we put
\[
n_k = N_k, \quad n_i = N_k - \sum_{j=i}^{k-1} N_j \quad (1 \leq i \leq k - 1),
\]
we have \( n_k > \cdots > n_1 \geq N_k/2 \geq 1 \), and
\[
n_j - n_i = \sum_{h=i}^{j-1} N_h \quad (1 \leq j < i \leq k),
\]
which is equivalent to (1).

**Lemma 2.** Let \( r \) be a positive integer, and for \( D \geq 1 \) let \( \mathcal{D} = \mathcal{D}(D, r) \) be the set of positive integers \( d \leq D \) of the form
\[
d = d_1 p, \quad (p, r) = 1, \quad q \mid d_1 \Rightarrow q \mid r,
\]
where \( p \) and \( q \) denote primes. Then we have
\[
\frac{1}{r} \sum_{d \in \mathcal{D}} \frac{1}{d} = \frac{\log \log(D + 2)}{\varphi(r)} + O(1)
\]
and
\[
d \left( \mathbb{N} \setminus \bigcup_{d \in \mathcal{D}} d(r \mathbb{N} - 1) \right) \ll \frac{\varphi(r)}{\log \log(D + 2)},
\]
where \( \varphi \) 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
\[
\frac{1}{d} = \frac{1}{d_1 \sum_{p \leq D} \frac{1}{p}} \leq \frac{1}{d_1} \sum_{p \leq D} \frac{1}{p} \leq \prod_{p \leq r} \left( 1 - \frac{1}{p} \right)^{-1} \sum_{p \leq D} \frac{1}{p} = \frac{r}{\varphi(r)} \left( \log \log(D + 2) + O(1) \right)
\]
and
\[
\frac{1}{d} \geq \frac{1}{d_1} \sum_{p \leq D} \frac{1}{p} \geq \left( \frac{r}{\varphi(r)} - \sum_{d_1 > \sqrt{D}} \frac{1}{d_1} \right) \log \log(D + 2) + O(r).
\]
This yields (3), since
\[
\sum_{d_1 > \sqrt{D}} \frac{1}{d_1} \ll D^{-1/4} \sum_{d_1 \ll \sqrt{D}} d_1^{-1/2} = D^{-1/4} \prod_{p \mid r} (1 - p^{-1/2})^{-1}
\]
\[
\ll D^{-1/4} \exp \left( 2 \sum_{p \mid r} p^{-1/2} \right) \ll rD^{-1/4}.
\]
For the proof of (4) we may suppose
\[
\log \log D \geq C \varphi(r),
\]
where \( C \) is an arbitrary, but fixed, positive constant. Set \( S = \bigcup_{d \in \mathcal{D}} d(r\mathbb{N} - 1) \) and define

\[
f(n) = \sum_{d|n, d \in \mathcal{D}, n/d = -1 \mod r} 1.
\]

Thus \( f(n) \geq 0 \) 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 \( d(\mathbb{N} \setminus S) \).

Putting

\[
M = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} f(n) = \sum_{d \in \mathcal{D}} \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x/d, n/d = -1 \mod r} 1 = \frac{1}{r} \sum_{d \in \mathcal{D}} \frac{1}{d},
\]

we have

\[
d(\mathbb{N} \setminus S) = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x, f(n)=0} 1 \leq \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x, f(n) \geq M/2} 1 \leq \frac{4}{M^2} \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} (f(n) - M)^2 = \frac{4}{M^2} \left( \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} f(n)^2 - M^2 \right)
\]

\[
= \frac{4}{M^2} \left( M_2 - M^2 \right), \quad \text{say}.
\]

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

\[
M_2 = \lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} f(n)^2 \leq \left( \log \log D \right)^2 \left( \frac{1}{\varphi(r)} \right) + O(\log \log D \varphi(r))
\]

with an absolute \( O \)-constant.

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

\[
M_2 = \sum_{d, d' \in \mathcal{D}} \frac{1}{[d, d']} \lim_{x \to \infty} \frac{[d, d']}{x} \sum_{n \leq x/[d, d']} 1,
\]

where (*) denotes the condition

\[
\frac{nd}{(d, d')} \equiv \frac{nd'}{(d, d')} \equiv -1 \mod r.
\]

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 \( d_1 = d'_1, p \equiv p' \mod r \), and in this case the limit in the last expression equals \( 1/r \). Thus we get

\[
M_2 \leq \frac{1}{r} \sum_{d_1 \leq D} \frac{1}{d_1} \sum_{p, p' \leq D, p \equiv p' \mod r} \left( \frac{1}{p, p'} \right)
\]

\[
\leq \frac{1}{\varphi(r)} \sum_{p \leq D} \frac{1}{p} \sum_{p' \leq D, p' \equiv p \mod r} \left( \frac{1}{p'} + O(1) \right).
\]

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
ON A CONJECTURE OF BALOG

The innermost sum equals

\[ \sum_{p' \leq p < D \atop p' \equiv p \mod r} \frac{1}{p'} + O \left( \sum_{\nu \leq \nu' \atop \nu \equiv \nu' \mod r} \frac{1}{\nu} \right) \]

\[ = \int_{\nu'}^{D} \pi(x, r, p) \frac{dx}{x^2} + O(1) = \frac{\log \log D}{\varphi(r)} + O(1), \]

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

\[ M_2 \leq \frac{1}{\varphi(r)} \left( \sum_{p \leq D} \frac{1}{p} \left( \frac{\log \log D}{\varphi(r)} + O(1) \right) \right) \]

\[ \leq \left( \frac{\log \log D}{\varphi(r)} \right)^2 + O \left( \frac{\log \log D}{\varphi(r)} \right), \]

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

4. Proof of the Theorem. For \( x > 0 \) let \( d_x(\cdot) \) be defined by

\[ d_x(M) = \frac{1}{x} \sum_{\nu \in [x]} 1 \quad (M \subseteq \mathbb{N}), \]

so that

\[ d(M) = \liminf_{x \to \infty} d_x(M), \quad d(M) = \limsup_{x \to \infty} d_x(M). \]

If \( \lambda \geq 1 \), then obviously

\[ d_x(\lambda M) = (1/\lambda) d_x(\lambda M) \leq d_x(M), \]

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

\[ d_x(M + t) = d_x(M) + o(1) \quad \text{as} \quad x \to \infty. \]

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

\[ B_{i, d} = n_i (A + d) \cap d\mathbb{N} \quad (1 \leq i \leq k, d \in \mathbb{N}), \]

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

\[ \left( \sum_{i=1}^k d_x(B_{i, d}) \geq \sum_{i=1}^k d_x(B_{i, d}) - \sum_{1 \leq i < j \leq k} d_x(B_{i, d} \cap B_{j, d}). \right. \]

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

\[ d_x(A_N \cap (A_N + 1)), \]

where \( N \geq \max(d, n) \), and bound the first term, averaged over a suitable range for \( d \), from below in terms of \( d_{x/n_k}(A) \). 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

\[ d_x(B_{i, d}) = d_x \left( n_i (A + d) \cap d\mathbb{N} \right) = \frac{1}{n_i} d_{x/n_i} \left( (A + d) \cap d\mathbb{N} \right) \]

\[ \geq \frac{1}{n_k} d_{x/n_k} (A \cap d\mathbb{N}) + o(1), \]
where \( T_i = (n/n_i)N - 1 \). Moreover, for \( 1 \leq i < j \leq k \), we get
\[
d_x(B_{i,d} \cap B_{j,d}) \leq d_x(n_i(A + d) \cap n_j(A + d) \cap d_n n_i N)
\]
\[
\leq d_x \left( \frac{n_i}{n_i} \frac{A \cap d N}{d} \cap \left( \frac{n_j}{n_j} \frac{A \cap d N}{d} + \frac{n_j - n_i}{n_i, n_j} \right) \right) + o(1)
\]
\[
\leq d_x(A_N \cap (A_N + 1)) + o(1),
\]
provided \( N \geq \max(n, d) \), where the last step follows from (1) and the definition of \( A_N \). Substituting these estimates together with the trivial bound
\[
d_x \left( \bigcup_{i=1}^k B_{i,d} \right) \leq d_x(d n \mathcal{N}) \leq \frac{1}{d n}
\]
into (8) yields
\[
\frac{1}{d n} \geq \frac{1}{n_k} \sum_{i=1}^k d_{x/n_k}(A \cap d T_i) - k^2 d_x(A_N \cap (A_N + 1)) + o(1)
\]
for every fixed \( d \in \mathbb{N} \) and \( N \geq \max(n, d) \).

We now fix \( \delta \geq 1 \) and let \( \mathcal{D} = \mathcal{D}(D, n/n_i) \) be defined as in Lemma 2, with \( r = n/n_i \). Since \( \mathcal{D}(D, r) \) depends only on the set of prime factors of \( r \), and the numbers \( n/n_i = \prod_{j=1}^k n_i^j / n_i, 1 \leq i \leq k \), have the same set of prime factors, this definition does not depend on the choice of the index \( i \). Summing the last inequality over \( d \in \mathcal{D} \), we obtain, for \( N \geq \max(n, D) \),
\[
\frac{1}{n} \sum_{d \in \mathcal{D}} \frac{1}{d} \geq \frac{1}{n_k} \sum_{i=1}^k \sum_{d \in \mathcal{D}} d_{x/n_k}(A \cap d T_i) - D k^2 d_x(A_N \cap (A_N + 1)) + o(1)
\]
\[
\geq \frac{1}{n_k} \sum_{i=1}^k d_{x/n_k}(A \cap S_i) - D k^2 d_x(A_N \cap (A_N + 1)) + o(1),
\]
where
\[
S_i = \bigcup_{d \in \mathcal{D}} d T_i = \bigcup_{d \in \mathcal{D}} \left( \frac{n}{n_i} N - 1 \right).
\]

Letting \( x \to \infty \), we deduce
\[
D k^2 d(A_N \cap (A_N + 1)) \geq \frac{1}{n_k} \sum_{i=1}^k d(A \cap S_i) - \frac{1}{n} \sum_{d \in \mathcal{D}} \frac{1}{d}
\]
\[
\geq \frac{k}{n_k} d(A) - \frac{1}{n_k} \sum_{i=1}^k d(N \setminus S_i) - \frac{1}{n} \sum_{d \in \mathcal{D}} \frac{1}{d}.
\]

By Lemma 2 we have
\[
\frac{n_k}{n} \sum_{d \in \mathcal{D}} \frac{1}{d} \ll \frac{\log \log(D + 2)}{\varphi(n/n_k)} + 1
\]
and
\[
d(N \setminus S_i) \ll \frac{\varphi(n/n_i)}{\log \log(D + 2)}.
\]
Since
\[ \frac{\varphi(n/n_{i})}{n/n_{i}} = \prod_{p | n/n_{i}} \left(1 - \frac{1}{p}\right) = \prod_{p | n_{1}, \ldots, n_{k}} \left(1 - \frac{1}{p}\right) \]
is independent of the choice of \( i \), and since, in view of (1),
\[ n_{1} \leq n_{i} \leq n_{k} \leq 2n_{1}, \]
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),
\[ Dk^{2}d(A_{N} \cap (A_{N} + 1)) \geq \frac{k}{n_{k}} \left( d(A) + O\left(\frac{1}{\sqrt{k}}\right) \right) \]
with an absolute \( O \)-constant. If now \( d(A) \geq \varepsilon (> 0) \), then by choosing \( k = k(\varepsilon) \) sufficiently large (which is possible by Lemma 1), the \( O \)-term becomes \( \leq \varepsilon/2 \), and we get
\[ d(A_{N} \cap (A_{N} + 1)) \geq \delta(\varepsilon) \quad (N \geq N(\varepsilon)), \]
with
\[ \delta(\varepsilon) = \frac{\varepsilon}{2D(k)kn_{k}}, \quad N(\varepsilon) = \max(n, D(k)), \]
as asserted in the Theorem.

By a minor modification of the proof, one can show that the theorem remains valid, when \( d \) is replaced by the upper density \( \bar{d} \).

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

Department of Mathematics, University of Illinois, Urbana, Illinois 61801

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