ACTA ARITHMETICA
162.3 (2014)

On the concentration of certain additive functions

by

DIMITRIS KOUKOULOPOULOS (Montréal)

In memoriam Jonas Kubilius

1. Introduction. An arithmetic function $f : \mathbb{N} \to \mathbb{R}$ is called *additive* if $f(mn) = f(m)+f(n)$ whenever $(m,n) = 1$. According to the Kubilius probabilistic model of the integers, statistical properties of additive functions can be modeled by statistical properties of sums of independent random variables. We describe this model in the case that f is a *strongly additive function*, that is, f satisfies the relation $f(n) = \sum_{p|n} f(p)$; the general case is slightly more involved. Let $\mathbb{P}$ denote the set of prime numbers and consider a sequence of independent Bernoulli random variables $\{X_p : p \in \mathbb{P}\}$ such that

$$\mathbf{Prob}(X_p = 1) = \frac{1}{p} \quad \text{and} \quad \mathbf{Prob}(X_p = 0) = 1 - \frac{1}{p}.$$

The random variable X_p can be thought of as a model of the characteristic function of the event $\{n \in \mathbb{N} : p \,|\, n\}$. Then a probabilistic model for f is given by the random variable $\sum_p f(p)X_p$.

The above model and well-known facts from probability theory lead to the prediction that the values of f follow a certain distribution, possibly after rescaling them appropriately. In fact, the Erdős–Wintner theorem [8] states that if the series

$$\sum_{|f(p)|\le 1} \frac{f(p)}{p}, \quad \sum_{|f(p)|\le 1} \frac{f^2(p)}{p}, \quad \sum_{|f(p)|>1} \frac{1}{p} \tag{1.1}$$

converge, then f has a limiting distribution, in the sense that there is a distribution function $F : \mathbb{R} \to [0,1]$ such that

$$F_x(u) := \frac{1}{\lfloor x \rfloor}\{n \le x : f(n) \le u\}| \to F(u) \quad \text{as } x \to \infty$$

2010 *Mathematics Subject Classification*: Primary 11N60, 11K65.
Key words and phrases: additive functions, Erdős–Wintner theorem, concentration of distribution.

DOI: 10.4064/aa162-3-2

for every $u \in \mathbb{R}$ that is a point of continuity of F; the characteristic function of F is given by

$$\hat{F}(\xi) = \prod_p \left\{ \left(1 - \frac{1}{p}\right) \sum_{k \geq 0} \frac{e^{i\xi f(p^k)}}{p^k} \right\}.$$

Conversely, if f has a limiting distribution, then the three series in (1.1) converge.

One way to measure the regularity of the distribution of the set $\{f(n) : n \in \mathbb{N}\}$ is by its concentration. In general, given a distribution function $G : \mathbb{R} \to [0,1]$, we define its *concentration function* to be

$$Q_G(\epsilon) = \sup_{u \in \mathbb{R}} \{G(u+\epsilon) - G(u)\}.$$

We seek estimates for $Q_{F_x}(\epsilon)$, or for $Q_F(\epsilon)$ if f has a limiting distribution. There are various such results in the literature, a historic account of which is given in [1]. The most general estimate on $Q_{F_x}(\epsilon)$ is due to Ruzsa [12]. Improving upon bounds due to Erdős [5] and Halász [9], he showed that

$$Q_{F_x}(1) \ll \max_{\lambda \in \mathbb{R}} \frac{1}{\sqrt{\lambda^2 + \sum_{p \leq x} \min\{1, (f(p) - \lambda \log p)^2\}/p}}. \tag{1.2}$$

This result is best possible, as can be seen by taking $f(n) = c \log n$ or $f(n) = \omega(n) = \sum_{p|n} 1$. However, both of these functions satisfy $f(p) \gg 1$. So a natural question is whether it is possible to improve upon (1.2) in the case that $f(p)$ decays to zero. Erdős and Kátai [7], building on earlier work of Tjan [14] and Erdős [6], showed the following result:

Theorem 1.1 (Erdős, Kátai [7]). *Let $f : \mathbb{N} \to \mathbb{R}$ be an additive function such that*

$$\sum_{p > t^A} \frac{|f(p)|}{p} \ll \frac{1}{t} \quad (t \geq 1), \qquad |f(p_1) - f(p_2)| \gg \frac{1}{p_2^B} \quad (p_1, p_2 \in \mathbb{P},\, p_1 < p_2),$$

for some constants A and B. Then

$$Q_F(\epsilon) \asymp_{A,B} \frac{1}{\log(1/\epsilon)} \qquad (0 < \epsilon \leq 1/2);$$

besides A and B, the implied constant depends on the implied constants in the assumptions of the theorem.

On the other hand, when $f(p) \ll 1/p^\delta$, $p \in \mathbb{P}$, for some $\delta > 0$, then (1.2) applied to f/ϵ yields an upper bound for $Q_F(\epsilon)$ that is never better than $1/\sqrt{\log\log(1/\epsilon)}$, as can be seen by taking $\lambda = 0$.

Also, Erdős and Kátai studied $Q_F(\epsilon)$ in the case that $f(p) = (\log p)^{-c}$, $p \in \mathbb{P}$, for some $c \geq 1$. They showed that

$$\begin{cases} \epsilon^{1/c} \ll_c Q_F(\epsilon) \ll_c \epsilon^{1/c} \log\log^2(1/\epsilon) & \text{if } c > 1, \\ \epsilon \ll Q_F(\epsilon) \ll \epsilon \log(1/\epsilon) \log\log^2(1/\epsilon) & \text{if } c = 1, \end{cases} \tag{1.3}$$

for $0 < \epsilon \leq 1/3$. Furthermore, they conjectured that, for every fixed $c > 1$, we have

$$Q_F(\epsilon) \asymp_c \epsilon^{1/c} \quad (0 < \epsilon \leq 1).$$

The conjecture of Erdős and Kátai was proven for c large enough by de la Bretèche and Tenenbaum in [1]:

THEOREM 1.2 (de la Bretèche, Tenenbaum [1]). *Let $c \geq 1$, and $f : \mathbb{N} \to \mathbb{R}$ be an additive function such that $|f(p)| \asymp (\log p)^{-c}$ for every $p \in \mathbb{P}$ and*

$$|f(p_1) - f(p_2)| \gg \frac{p_2 - p_1}{p_2(\log p_2)^{c+1}} \quad (p_1, p_2 \in \mathbb{P},\ p_1 < p_2).$$

If c is large enough, then

$$Q_F(\epsilon) \asymp \epsilon^{1/c} \quad (0 < \epsilon \leq 1);$$

the implied constant depends at most on the implied constants in the assumptions of the theorem.

De la Bretèche and Tenenbaum derived their theorem from a general upper bound on $Q_F(\epsilon)$ that they showed when the sequence of prime values of f satisfies certain regularity assumptions. Their method uses a result from the theory of functions of *bounded mean oscillation*, first introduced by Diamond and Rhoads [2] in this context to study the concentration of $f(n) = \log(\phi(n)/n)$.

In this paper we give a proof of the full Erdős–Kátai conjecture using a more elementary method, similar to the ones in [6, 7]:

THEOREM 1.3. *Let $c \geq 1$, and $f : \mathbb{N} \to \mathbb{R}$ be an additive function with $f(p) = (\log p)^{-c}$ for all $p \in \mathbb{P}$. For $0 < \epsilon \leq 1/2$ we have*

$$\epsilon^{1/c} \ll Q_F(\epsilon) \ll \min\left\{\frac{c}{c-1}, \log\frac{1}{\epsilon}\right\}\epsilon^{1/c}.$$

REMARK 1.4. When $0 < c < 1$, the behavior of Q_F for f as in Theorem 1.3 is different. As Gérald Tenenbaum has pointed out to us in a private communication, in this case

$$Q_F(\epsilon) \asymp_c \epsilon \quad (0 < \epsilon \leq 1). \tag{1.4}$$

Corollary 1.3 and relation (1.4) give the concentration of an additive function

f with $f(p) = (\log p)^{-c}$, $p \in \mathbb{P}$, for all positive values of c except for $c = 1$, which is the only case remaining open.

We will prove Theorem 1.3 in Section 2. The method of its proof is quite flexible; in particular, it leads to a strengthening of Theorems 1.1 and 1.2. We phrase our more general result in terms of the distribution function

$$\mathcal{F}_y(u) = \prod_{p \le y}\left(1 - \frac{1}{p}\right) \sum_{\substack{p \mid n \Rightarrow p \le y \\ f(n) \le u}} \frac{1}{n} \quad (u \in \mathbb{R}),$$

defined for every $y \ge 1$. From a technical point of view, this function is more natural to work with than F_x. Indeed, a calculation of the characteristic function of $\mathcal{F}_y$ immediately implies that $\mathcal{F}_y$ converges to F weakly, provided that the latter is well-defined. It is relatively easy to pass from estimates for $Q_{\mathcal{F}_y}(\epsilon)$ to estimates for $Q_{F_x}(\epsilon)$.

With this notation, we have the following result (observe that by letting $y \to \infty$ in it, we deduce as special cases ([1]) Theorems 1.1 and 1.2):

Theorem 1.5. *Consider an additive function $f : \mathbb{N} \to \mathbb{R}$ for which there is a set of primes $\mathcal{P}$ and a constant $c \in [1, 2]$ such that*

$$|f(p)| \ll \frac{1}{(\log p)^c} \quad (p \in \mathcal{P}), \quad \textit{and} \quad \sum_{p \in \mathbb{P} \setminus \mathcal{P}} \frac{1}{p} \ll 1.$$

For $t \ge 2$ set

$$g(t) = \frac{\sup\{|f(p)|(\log p)^c : p \ge t,\, p \in \mathcal{P}\}}{(\log t)^c}$$

and assume that there is some $A \ge 1$ such that

$$|f(p_2) - f(p_1)| \gg \min\left\{\frac{g(p_2)(p_2 - p_1)}{p_2 \log p_2}, g(p_2^A)\right\} \quad (p_1, p_2 \in \mathcal{P},\, p_1 < p_2).$$

Then for $0 < \epsilon \le 1/2$ and $y \ge K(\epsilon)$, where $K(\epsilon) = \min\{p \in \mathbb{P} : g(p) \le \epsilon\}$, we have

$$\frac{1}{\log K(\epsilon)} \ll Q_{\mathcal{F}_y}(\epsilon) \ll_A \frac{\min\{1/(c-1), \log(1/\epsilon)\}}{\log K(\epsilon)};$$

besides A, the implied constant depends on the implied constants in the assumptions of the theorem.

As an immediate corollary, we deduce the following simpler to state result.

Corollary 1.6. *Let $f : \mathbb{N} \to \mathbb{R}$ be an additive function for which there is a constant $c \in [1, 2]$ such that the sequence $\{|f(p)|(\log p)^c : p$ prime$\}$*

([1]) To deduce Theorem 1.1, take $\mathcal{P} = \{p \in \mathbb{P} : |f(p)| \le p^{-1/(2A)}\}$.

is decreasing. Then for $0 < \epsilon \leq 1/2$ *and* $y \geq K(\epsilon)$, *where* $K(\epsilon) = \min\{p \in \mathbb{P} : |f(p)| \leq \epsilon\}$, *we have*

$$\frac{1}{\log K(\epsilon)} \ll Q_{\mathcal{F}_y}(\epsilon) \ll \frac{\min\{1/(c-1), \log(1/\epsilon)\}}{\log K(\epsilon)}.$$

Proof. If g is as in the statement of Theorem 1.5 with $\mathcal{P} = \mathbb{P}$, then we see immediately that $g(p) = |f(p)|$ for all $p \in \mathbb{P}$. Moreover, if $p_1 < p_2$ are two primes, then

$$\begin{aligned}\frac{|f(p_2) - f(p_1)|}{|f(p_2)|} &\geq \frac{|f(p_1)|}{|f(p_2)|} - 1 \geq \frac{(\log p_2)^c}{(\log p_1)^c} - 1 = \frac{(\log p_2)^c - (\log p_1)^c}{(\log p_1)^c} \\ &\geq \frac{c(p_2 - p_1)(\log p_1)^{c-1}}{p_2(\log p_1)^c} \geq \frac{p_2 - p_1}{p_2 \log p_2},\end{aligned}$$

by our assumption that $\{|f(p)|(\log p)^c : p \text{ prime}\}$ is decreasing and the Mean Value Theorem. So Theorem 1.5 can be applied and the claimed result follows. ∎

The lower bound in Theorem 1.5, which will be proven in Section 3, is a straightforward application of Theorem 1.2 in [1]. On the other hand, for the proof of the upper bound in Theorem 1.5, which will be given in Section 4, we use a combination of ideas from [6, 7]. Even though Theorem 1.3 is an immediate corollary of Theorem 1.5, applied with $\min\{c, 2\}$ in place of c, we have chosen to give the proof of both of them in full detail, so as to motivate certain choices in the proof of Theorem 1.5, which is rather technical.

A heuristic argument. There is a simple heuristic argument which motivates Theorem 1.5. We demonstrate it in the simpler setting of Corollary 1.6, that is, when the sequence $\{|f(p)|(\log p)^c : p \in \mathbb{P}\}$ is decreasing. For every integer n, we have

$$\sum_{p|n,\, p \geq K(\epsilon)} |f(p)| \leq \epsilon \sum_{p|n,\, p \geq K(\epsilon)} \frac{(\log K(\epsilon))^c}{(\log p)^c}.$$

Since for a typical integer n the sequence $\{\log\log p : p \mid n\}$ is distributed like an arithmetic progression of step 1 (see, for example, [10, Chapter 1]), we find that (2)

$$\sum_{p|n,\, p \geq K(\epsilon)} |f(p)| \lesssim \epsilon \sum_{j \geq \log\log K(\epsilon)} \frac{(\log K(\epsilon))^c}{e^{cj}} \ll \epsilon.$$

So only the prime divisors of n lying in $[1, K(\epsilon))$ are important for the size of $Q_F(\epsilon)$. Note that for a prime number $p < K(\epsilon)$, we have $|f(p)| > \epsilon$. Therefore if a and b are composed of primes within $[1, K(\epsilon))$, then it is reasonable

(2) The symbol '$\lesssim$' here is used in a non-rigorous fashion to denote 'roughly less than'. Similarly, the symbol '$\approx$' means 'roughly equal to'.

to expect that $|f(a) - f(b)|$ is large compared to ϵ, unless a and b have a large common factor. This leads to the prediction that $Q_F(\epsilon) \approx 1/\log K(\epsilon)$, which is confirmed by Theorem 1.5 when $c > 1$. However, when $c < 1$ this heuristic fails, as (1.4) shows, and the underlying reason is combinatorial: the pigeonhole principle implies the lower bound $Q_F(\epsilon) \gg_F \epsilon$ for the concentration function of any distribution function F (see also [7, Remark 1, p. 297]).

Notation. For an integer n we denote by $P^+(n)$ and $P^-(n)$ its largest and smallest prime factors, respectively, with the notational convention that $P^+(1) = 1$ and $P^-(1) = \infty$. The symbols p and p' always denote prime numbers. The set of all prime numbers is denoted by $\mathbb{P}$. Finally, given $\mathcal{P} \subset \mathbb{P}$ and real numbers $1 \le z \le w$, we write $\mathcal{P}(z, w)$ for the set of integers all of whose prime factors belong to $\mathcal{P} \cap (z, w]$.

2. The conjecture of Erdős and Kátai

Proof of Theorem 1.3. The lower bound follows by relation (1.3), with the implied constant depending on c. To remove this dependence, see Theorem 3.1 below.

It remains to show the corresponding upper bound. Before delving into the details of the proof, we give a brief outline of the main idea. For $\delta > 0$, we set $P_\delta = \exp(\delta^{-1/c})$, so $f(p) = \delta$ if and only if $p = P_\delta$. As the heuristic argument presented towards the end of Section 1 indicates, it suffices to bound $Q_{\mathcal{F}_q}(\epsilon)$, where $q := P_{2\epsilon}$. We split the elements of the set $\mathcal{M} := \{n \in \mathbb{P}(1, q) : u < f(n) \le u + \epsilon\}$ into subsets $\mathcal{M}_\delta := \{n \in \mathcal{M} : P_{2\delta} < P^-(n) \le P_\delta\}$, where $\delta \in \{2^j\epsilon : 1 \le j \le j_0\}$ with $j_0 = \lfloor(\log(1/\epsilon) - c\log\log 2)/\log 2\rfloor$. Then we find that

$$\sum_{n \in \mathcal{M}_\delta} \frac{1}{n} \approx \sum_{m \in \mathbb{P}(P_\delta, q)} \frac{1}{m} \sum_{\substack{P_{2\delta} < p \le P_\delta \\ u - f(m) < f(p) \le u - f(m) + \epsilon}} \frac{1}{p}.$$

Fix m for the moment and set $v_m = u - f(m)$. Then the variable p lies in the interval $I_m = [P_{\min\{v_m+\epsilon, 2\delta\}}, P_{\max\{v_m, \delta\}}]$, which is non-empty only when $v_m + \epsilon \ge \delta \ge v_m/2$. Since $\delta \ge 2\epsilon$ by assumption, we find that $v_m \asymp \delta \gg \epsilon$. So the interval I_m has double-logarithmic length ([3]) $\ll \epsilon/v_m \asymp \epsilon/\delta$. Hence the Prime Number Theorem [13, Theorem 1, p. 167] implies that $\sum_{p \in I_m} 1/p \lesssim \epsilon/\delta$, provided that I is not too short. Assuming that this is indeed the case, we deduce that

$$\sum_{n \in \mathcal{M}_\delta} \frac{1}{n} \lesssim \frac{\epsilon}{\delta} \sum_{m \in \mathbb{P}(P_\delta, q)} \frac{1}{m} \ll \frac{\epsilon \log q}{\delta \log P_\delta} \asymp \frac{\epsilon^{1-1/c}}{\delta^{1-1/c}}.$$

([3]) Given an interval $I = [\alpha, \beta]$, its *double-logarithmic length* is $\log\log\beta - \log\log\alpha$.

Summing the above inequality over $\delta \in \{2^j\epsilon : 1 \le j \le j_0\}$ implies that

$$Q_F(\epsilon) \approx Q_{\mathcal{F}_q}(\epsilon) \lesssim \frac{1}{\log q} \sum_{1 \le j \le j_0} \sum_{n \in \mathcal{M}_{2^j\epsilon}} \frac{1}{n} \lesssim \epsilon^{1/c} \sum_{1 \le j \le j_0} \frac{1}{2^{j(1-1/c)}}$$
$$\ll \min\left\{\frac{c}{c-1}, \log\frac{1}{\epsilon}\right\}\epsilon^{1/c},$$

which shows (heuristically at least) the desired result.

The main technical difficulty we have to surpass in order to make the above argument work is that the estimate $\sum_{p \in I_m} 1/p \ll \epsilon/\delta$, which we used above, might not be accurate for large δ (i.e. when P_δ is small). So below we shall employ a variation of the argument of this paragraph where, instead of looking where $P^-(n)$ lies, we will look where $\min\{p \,|\, n : p \ge q'\}$ lies, with q' being some small parameter chosen appropriately.

Without loss of generality, we may assume that $\epsilon \le 1/100^c$; otherwise (2.2) follows immediately by the trivial bound $Q_F(\epsilon) \le 1$. Define η by $P_\eta/\eta^2 = 1/\epsilon^2$. Note that $4\epsilon \le \eta \le 1/2$, since $P_{4\epsilon}/(4\epsilon)^2 \ge 1/\epsilon^2 \ge P_{1/2}/(1/2)^2$ by our assumption that $\epsilon \le 1/100^c$. Before we proceed further, we will prove that, for $v \in \mathbb{R}$, $\delta \in [2\epsilon, 1]$ and $2 \le z \le P_\delta$, we have

$$\sum_{\substack{z < p \le P_\delta \\ v < f(p) \le v+\epsilon}} \frac{1}{p} \ll \frac{\epsilon}{\delta} + \frac{1}{\sqrt{z}} \ll \begin{cases} \epsilon/\delta & \text{if } z \ge P_{2\delta} \ge P_\eta, \\ \epsilon/\delta + 1/(\log z)^2 & \text{otherwise.} \end{cases} \tag{2.1}$$

First, note that it suffices to show the first inequality. Indeed, if $z \ge P_{2\delta} \ge P_\eta$, then $P_{2\delta}/(2\delta)^2 \ge P_\eta/\eta^2 = 1/\epsilon^2$, and thus $\sqrt{z} \ge \sqrt{P_{2\delta}} \ge 2\delta/\epsilon$, which proves the second inequality of (2.1). Returning to the first inequality of (2.1), observe that the primes p on the left-hand side of (2.1) lie in the interval $[\max\{z, P_{v+\epsilon}\}, P_{\max\{v,\delta\}}] =: [\alpha, \beta]$. For this interval to be non-empty we need that $v + \epsilon \ge \delta$. Since $\delta \ge 2\epsilon$, we deduce that $v \ge \epsilon$ and thus $2v \ge v + \epsilon \ge \delta$. So we have

$$\log\frac{\log\beta}{\log\alpha} \le \log\frac{\log P_v}{\log P_{v+\epsilon}} = \frac{1}{c}\log\frac{v+\epsilon}{v} \le \frac{\epsilon}{v} \le \frac{2\epsilon}{\delta}.$$

Thus, if $\beta \ge \alpha + \sqrt{\alpha}$, then covering the interval $[\alpha, \beta]$ by subintervals of the form $[y, y + \sqrt{y})$ and applying the Brun–Titchmarsch inequality [13, Theorem 9, p. 73] to each one of them yields

$$\sum_{\substack{z < p \le P_\delta \\ v < f(p) \le v+\epsilon}} \frac{1}{p} \ll \frac{\epsilon}{\delta},$$

which proves (2.1). Finally, if $\beta < \alpha + \sqrt{\alpha}$, then

$$\sum_{\substack{z<p\le P_\delta\\ v<f(p)\le v+\epsilon}} \frac{1}{p} \le \sum_{\alpha\le p\le\beta} \frac{1}{p} \le \frac{\beta-\alpha+2}{\alpha} \ll \frac{1}{\sqrt{\alpha}} \le \frac{1}{\sqrt{z}}$$

and (2.1) follows in this last case too.

We are now ready to show the upper bound implicit in Theorem 1.3. Fix for the moment $y \ge q = P_{2\epsilon}$ and $u \in \mathbb{R}$. Given $n \in \mathbb{P}(1,y)$ with $u < f(n) \le u+\epsilon$, we write $n = ab$, where a is square-free, b is square-full and $(a,b) = 1$. We further decompose $a = a_1a_2$, where $P^+(a_1) \le q < P^-(a_2)$. Therefore

$$\begin{aligned}\sum_{\substack{P^+(n)\le y\\ u<f(n)\le u+\epsilon}} \frac{1}{n} &= \sum_{\substack{P^+(b)\le y\\ b\,\text{square-full}}} \frac{1}{b} \sum_{\substack{a_2\in\mathbb{P}(q,y)\\ (a_2,b)=1}} \frac{\mu^2(a_2)}{a_2} \sum_{\substack{P^+(a_1)\le q,\,(a_1,b)=1\\ u-f(a_2b)<f(a_1)\le u-f(a_2b)+\epsilon}} \frac{\mu^2(a_1)}{a_1}\\ &\le \sum_{\substack{P^+(b)\le y\\ b\,\text{square-full}}} \frac{1}{b} \sum_{a_2\in\mathbb{P}(q,y)} \frac{1}{a_2} \sup_{v\in\mathbb{R}} \Big\{ \sum_{\substack{P^+(a_1)\le q\\ v<f(a_1)\le v+\epsilon}} \frac{\mu^2(a_1)}{a_1} \Big\}\\ &\ll \frac{\log y}{\log q} \sup_{v\in\mathbb{R}} \Big\{ \sum_{\substack{P^+(a_1)\le q\\ v<f(a_1)\le v+\epsilon}} \frac{\mu^2(a_1)}{a_1} \Big\}.\end{aligned}$$

Since $\log q \asymp \epsilon^{-1/c}$, Theorem 1.3 will follow from the estimate

$$\sum_{\substack{P^+(n)\le q\\ v<f(n)\le v+\epsilon}} \frac{\mu^2(n)}{n} \ll \min\Big\{\frac{c}{c-1}, \log\frac{1}{\epsilon}\Big\} \quad (v\in\mathbb{R}) \tag{2.2}$$

by letting $y \to \infty$. Set $J = -1 + \lfloor \log(\eta/\epsilon)/\log 2 \rfloor \in \mathbb{N}$ and, for $j \ge 0$, define $q_j = P_{2^{j+1}\epsilon}$, so that $q = q_0 > \cdots > q_J \ge P_\eta$. Fix $v \in \mathbb{R}$ and let

$$\mathcal{N} = \{n \in \mathbb{N} : \mu^2(n) = 1,\ P^+(n) \le q_0,\ v < f(n) \le v+\epsilon\}.$$

As in the heuristic argument of the first section, we partition $\mathcal{N}$ into certain subsets and estimate the contribution of each one of them to $\sum_{n\in\mathcal{N}} 1/n$ separately. The difference is that instead of looking at the location of $P^-(n)$, we write $n = an'$ with $P^+(a) \le q_J < P^-(n')$ and look at the location of $p = P^-(n')$. An additional fact that we shall take advantage of is that if $n' = pb$ and $\log p \asymp \log P_\delta$, then, for fixed b, the number $f(a)$ lies in an interval of length $\ll \epsilon + f(p) \ll \epsilon + \delta \ll \delta$, which allows us to gain an additional crucial saving in our estimate for $\sum_{n\in\mathcal{N}} 1/n$. So we write $\mathcal{N} = \bigcup_{j=0}^J \mathcal{N}_j$, where $\mathcal{N}_0 = \{n \in \mathcal{N} : P^+(n) \le q_J\}$ and

$$\mathcal{N}_j = \{n \in \mathcal{N} : n = apb,\ P^+(a) \le q_J < p < P^-(b),\ q_j < p \le q_{j-1}\}$$

for $j \in \{1,\ldots,J\}$.

First, we bound $\sum_{n\in\mathcal{N}_0} 1/n$. If $n>1$, then we write $n=mP^+(n)=mp'$. Thus

$$(2.3)\quad \sum_{n\in\mathcal{N}_0}\frac{1}{n}\le 1+\sum_{P^+(m)\le q_J}\frac{1}{m}\sum_{\substack{P^+(m)<p'\le q_J\\ v-f(m)<f(p')\le v-f(m)+\epsilon}}\frac{1}{p'}$$
$$\ll 1+\sum_{P^+(m)\le q_J}\frac{1}{m}\left(\frac{\epsilon}{2^J\epsilon}+\frac{1}{\log^2(1+P^+(m))}\right)\ll 1+\frac{\log q_J}{2^J}$$

by (2.1).

Next, we bound $\sum_{n\in\mathcal{N}_j} 1/n$ for $j\in\{1,\ldots,J\}$. We have

$$(2.4)\quad \sum_{n\in\mathcal{N}_j}\frac{1}{n}\le\sum_{b\in\mathbb{P}(q_j,q_0)}\frac{1}{b}\sum_{P^+(a)\le q_J}\frac{\mu^2(a)}{a}\sum_{\substack{q_j<p\le q_{j-1}\\ v-f(a)-f(b)<f(p)\le v-f(a)-f(b)+\epsilon}}\frac{1}{p}$$
$$\le\sum_{b\in\mathbb{P}(q_j,q_0)}\frac{1}{b}\sup_{w\in\mathbb{R}}\Bigg\{\sum_{P^+(a)\le q_J}\frac{\mu^2(a)}{a}\sum_{\substack{q_j<p\le q_{j-1}\\ w-f(a)<f(p)\le w-f(a)+\epsilon}}\frac{1}{p}\Bigg\}$$
$$\ll\frac{\log q_0}{\log q_j}\sup_{w\in\mathbb{R}}\Bigg\{\sum_{P^+(a)\le q_J}\frac{\mu^2(a)}{a}\sum_{\substack{q_j<p\le q_{j-1}\\ w-f(a)<f(p)\le w-f(a)+\epsilon}}\frac{1}{p}\Bigg\}.$$

Fix some $w\in\mathbb{R}$ and consider a with $P^+(a)\le q_J$ and $p\in(q_j,q_{j-1}]$ with $w<f(a)+f(p)\le w+\epsilon$, as above. Since $|f(p)|\le 2^{j+1}\epsilon$ for $p>q_j$, we must have $|f(a)-w|<2^{j+2}\epsilon$. So

$$(2.5)\quad \sum_{P^+(a)\le q_J}\frac{\mu^2(a)}{a}\sum_{\substack{q_j<p\le q_{j-1}\\ w-f(a)<f(p)\le w-f(a)+\epsilon}}\frac{1}{p}$$
$$=\sum_{\substack{P^+(a)\le q_J\\ |f(a)-w|<2^{j+2}\epsilon}}\frac{\mu^2(a)}{a}\sum_{\substack{q_j<p\le q_{j-1}\\ w-f(a)<f(p)\le w-f(a)+\epsilon}}\frac{1}{p}\ll\frac{\epsilon}{2^j\epsilon}\sum_{\substack{P^+(a)\le q_J\\ |f(a)-w|<2^{j+2}\epsilon}}\frac{\mu^2(a)}{a},$$

by the first part of (2.1) applied with $w-f(a)$, $2^j\epsilon$ and q_j in place of v, δ and z, respectively, since $q_j\ge q_J\ge P_\eta$. Finally, if $a>1$, then we write $a=mP^+(a)=mp'$. So we find that

$$\sum_{\substack{P^+(a)\le q_J\\ |f(a)-w|<2^{j+2}\epsilon}}\frac{\mu^2(a)}{a}\le 1+\sum_{P^+(m)\le q_J}\frac{1}{m}\sum_{\substack{P^+(m)<p'\le q_J\\ |f(p')-(w-f(m))|<2^{j+2}\epsilon}}\frac{1}{p'}.$$

For every fixed $m \in \mathbb{N}$, we have

$$\sum_{\substack{P^+(m)<p'\le q_J\\ |f(p')-(w-f(m))|<2^{j+2}\epsilon}} \frac{1}{p'} \ll \frac{2^j\epsilon}{2^J\epsilon} + \frac{1}{\log^2(1+P^+(m))},$$

by the second part of (2.1) with $2^j\epsilon$, $2^{J+1}\epsilon$ and $P^+(m)$ in place of ϵ, δ and z, respectively ([4]), and with $v \in \{w - f(m) + h \cdot 2^j\epsilon : h \in [-4,4) \cap \mathbb{Z}\}$. So we find that

$$\sum_{\substack{P^+(a)\le q_J\\ |f(a)-w|<2^{j+1}\epsilon}} \frac{\mu^2(a)}{a} \ll 1 + \sum_{P^+(m)\le q_J} \frac{1}{m}\left(\frac{2^j\epsilon}{2^J\epsilon} + \frac{1}{\log^2(1+P^+(m))}\right) \ll 1 + \frac{\log q_J}{2^{J-j}}.$$

Combining the above inequality with (2.4) and (2.5), we deduce that

$$\sum_{n\in\mathcal{N}_j} \frac{1}{n} \ll \frac{1}{2^j}\left(1 + \frac{\log q_J}{2^{J-j}}\right)\frac{\log q_0}{\log q_j} \ll \frac{1}{2^{j(1-1/c)}}\left(1 + \frac{\log q_J}{2^{J-j}}\right).$$

Together with (2.3), this implies that

$$\begin{aligned}\sum_{n\in\mathcal{N}} \frac{1}{n} &\ll \sum_{j=0}^{J} \frac{1}{2^{j(1-1/c)}}\left(1 + \frac{\log q_J}{2^{J-j}}\right) = \sum_{j=0}^{J} \frac{1}{2^{j(1-1/c)}} + \sum_{j=0}^{J} \frac{2^{j/c}\log q_J}{2^J}\\ &\ll \min\left\{\frac{c}{c-1}, J\right\} + \frac{\min\{c,J\}\log q_J}{2^{J(1-1/c)}}.\end{aligned}$$

Furthermore, $2^J \asymp \eta/\epsilon$ and, as a result,

$$\log q_J = (2^{J+1}\epsilon)^{-1/c} \asymp \eta^{-1/c} = \log P_\eta = 2\log(\eta/\epsilon) \ll J \ll \log(1/\epsilon).$$

Thus

$$\begin{aligned}\sum_{n\in\mathcal{N}} \frac{1}{n} &\ll \min\left\{\frac{c}{c-1}, J\right\} + \frac{J\min\{c,J\}}{2^{J(1-1/c)}} \ll \min\left\{\frac{c}{c-1}, J\right\}\\ &\ll \min\left\{\frac{c}{c-1}, \log\frac{1}{\epsilon}\right\},\end{aligned}$$

by the inequality $2^{J(1-1/c)} \gg J^2$ if $c \ge 2$ and the inequality $2^{J(1-1/c)} \gg \max\{1, J(1-1/c)\}$ if $1 \le c \le 2$. Therefore (2.2) follows, thus completing the proof of the theorem. ∎

3. The lower bound in Theorem 1.5. In this section we derive the lower bound in Theorem 1.5 from the following general result, which is a corollary of Theorem 1.2 in [1].

([4]) Note that the parameter η is not involved in the second part, so the same proof allows us to replace ϵ with $2^j\epsilon$.

THEOREM 3.1. *Let $f:\mathbb{N}\to\mathbb{R}$ be an additive function and $0<\epsilon<1$. If there is a set of primes $\mathcal{P}$ and some $M\ge 2$ such that*

$$\sum_{p\in\mathbb{P}\setminus\mathcal{P}}\frac{1}{p}\ll 1 \quad \textit{and} \quad \sum_{p\in\mathcal{P},\, p>M}\frac{|f(p)|}{p}\ll\epsilon,$$

then, for $y\ge M$, we have

$$Q_{\mathcal{F}_y}(\epsilon)\gg\frac{1}{\log M};$$

the implied constant depends at most on the implied constants implicit in the assumptions of the theorem.

Proof. Let M_0 be a large constant to be chosen later. If $M\le y\le M_0$, then the theorem follows by the trivial bound $Q_{\mathcal{F}_y}(\epsilon)\gg 1/\log y$, which holds since 1 is always in $\{n\in\mathbb{N}: P^+(n)\le y,\, |f(n)|<\epsilon/2\}$. Assume now that $y\ge M_0$ and set $M'=\max\{M,M_0\}$, so that $y\ge M'$. Let

$$C=\frac{1}{\epsilon}\sum_{p\in\mathcal{P},\, p>M'}\frac{|f(p)|}{p}\ll 1.$$

Define $g:\mathbb{N}\to\mathbb{R}$ by

$$g(n)=\begin{cases} f(n) & \text{if } n\in\mathcal{P}(1,y),\\ 0 & \text{otherwise,}\end{cases}$$

and call G its distribution function. Then Theorem 1.2 in [1] yields ([5])

$$Q_G(3C\epsilon)\ge\left(1-2\frac{C\epsilon}{3C\epsilon}+o_{M'\to\infty}(1)\right)\prod_{p\le M'}\left(1-\frac{1}{p}\right)\gg\frac{1}{\log M'}.$$

So, by the pigeonhole principle, we deduce that

$$Q_G(\epsilon)\ge\frac{Q_G(3C\epsilon)}{3C+1}\gg_C\frac{1}{\log M'}\asymp\frac{1}{\log M},\tag{3.1}$$

provided that M_0 is large enough. Finally,

$$\begin{aligned} Q_G(\epsilon)&=\sup_{u\in\mathbb{R}}\Big\{\prod_{p\in\mathcal{P}\cap[1,y]}\left(1-\frac{1}{p}\right)\sum_{\substack{n\in\mathcal{P}(1,y)\\ u<f(n)\le u+\epsilon}}\frac{1}{n}\Big\}\\ &\le Q_{\mathcal{F}_y}(\epsilon)\prod_{p\in\mathbb{P}\setminus\mathcal{P}}\left(1-\frac{1}{p}\right)^{-1}\ll Q_{\mathcal{F}_y}(\epsilon),\end{aligned}$$

which together with (3.1) completes the proof of the theorem. ∎

Proof of the lower bound in Theorem 1.5. The definition of g implies that the function $t\mapsto g(t)(\log t)^c$ is decreasing. Thus, for $p\in\mathcal{P}$ with $p\ge K(\epsilon)$,

([5]) In [1, Theorem 1.2], the authors let $\epsilon\to 0$. However, an easy modification of their proof allows us to let instead $M'\to\infty$.

we have

$$|f(p)| \le g(p) \le \frac{g(K(\epsilon))(\log K(\epsilon))^c}{(\log p)^c} \le \frac{\epsilon(\log K(\epsilon))^c}{(\log p)^c}.$$

Consequently,

$$\sum_{p>K(\epsilon)} \frac{|f(p)|}{p} \ll \epsilon,$$

which implies that the hypotheses of Theorem 3.1 are satisfied with $M = K(\epsilon)$ and $\mathcal{P}$, and the desired lower bound follows. ∎

4. The upper bound in Theorem 1.5. We conclude the paper by showing the upper bound in Theorem 1.5. We start with the following technical lemma whose hypotheses mimic all the crucial facts about the additive function $f(n) = \sum_{p|n} (\log p)^{-c}$ that we used in the proof of Theorem 1.3.

LEMMA 4.1. *Let $f : \mathbb{N} \to \mathbb{R}$ be an additive function for which there is a set of primes $\mathcal{P}$ and a decreasing function $P_f : (0, 1] \to [2, \infty)$ such that*

$$\sum_{p\in\mathbb{P}\setminus\mathcal{P}} \frac{1}{p} \ll 1, \tag{4.1}$$

$$|f(p)| \le \epsilon \quad (0 < \epsilon \le 1,\ p \in \mathcal{P},\ p > P_f(\epsilon)). \tag{4.2}$$

Furthermore, assume that there are $\lambda \in (0, 1]$ and $\rho \ge 1$ such that

$$\sum_{\substack{p\in\mathcal{P}\cap(z,w] \\ u<f(p)\le u+\epsilon}} \frac{1}{p} \ll \begin{cases} \epsilon/\delta & \text{if } z \ge P_f(2\delta) \ge (2\delta/\epsilon)^\rho, \\ \epsilon/\delta + 1/(\log z)^2 & \text{otherwise,} \end{cases} \tag{4.3}$$

for all $u \in \mathbb{R}$, $0 < \epsilon \le \delta \le 1$ and $2 \le z \le w \le \min\{P_f(\delta), P_f(\epsilon)^\lambda\}$. Let $0 < \epsilon \le \delta \le 1$ be such that $P_f(\delta) \le P_f(\epsilon)^\lambda$, set $q_j = P_f(2^j\delta)$ for $j \ge 0$, and consider $J \in \{0\} \cup \{j \in \mathbb{N} : 2^j \le 1/\delta \text{ and } q_j \ge (2^j\delta/\epsilon)^\rho\}$. For $y \ge q_0$, we then have

$$Q_{\mathcal{F}_y}(\epsilon) \ll \frac{1}{\log q_0} + \frac{\epsilon}{\lambda\delta} \sum_{j=1}^{J} \frac{1}{2^j \log q_j} + \frac{\epsilon}{\delta} \sum_{j=0}^{J} \frac{\log q_J}{2^J \log q_j}; \tag{4.4}$$

the implied constant depends at most on the implied constants in (4.1) *and* (4.3).

REMARK 4.2. The parameters λ and ρ, and the set $\mathcal{P}$ are introduced to make Lemma 4.1 more applicable. One can think of P_f defined by $P_f(\epsilon) = \max\{p \in \mathbb{P} : |f(p)| > \epsilon\}$. Condition (4.3) can be motivated as follows. Assume that

$$\sum_{p>P_f(\alpha)} \frac{|f(p)|}{p} \ll \alpha \quad (0 < \alpha \le 1). \tag{4.5}$$

We have $|f(p)| \approx \delta$ for $p \in (P_f(2\delta), P_f(\delta)]$. So if the sequence $\{f(p) : p \in \mathbb{P}\}$ is 'well-spaced', then we expect that

$$\sum_{\substack{P_f(2\delta)<p\le P_f(\delta)\\ w<f(p)\le w+\epsilon}} \frac{1}{p} \lesssim \frac{\epsilon}{\delta} \sum_{P_f(2\delta)<p\le P_f(\delta)} \frac{1}{p} \approx \frac{\epsilon}{\delta^2} \sum_{P_f(2\delta)<p\le P_f(\delta)} \frac{|f(p)|}{p} \ll \frac{\epsilon}{\delta},$$

by (4.5).

Proof of Lemma 4.1. Fix for the moment $u \in \mathbb{R}$. Given $n \in \mathbb{P}(1, y)$ with $u < f(n) \le u + \epsilon$, we write $n = ab$, where a is square-free, b is square-full and $(a, b) = 1$. We further decompose $a = a_1a_2a_3$, where $a_1 \in \mathcal{P}(1, q_0)$, $a_2 \in \mathcal{P}(q_0, y)$ and all prime factors of a_3 lie in $\mathcal{Q} := \mathbb{P} \setminus \mathcal{P}$. So

$$\tag{4.6} \begin{aligned} \sum_{\substack{P^+(n)\le y\\ u<f(n)\le u+\epsilon}} \frac{1}{n} &= \sum_{\substack{P^+(b)\le y\\ b\,\text{square-full}}} \frac{1}{b} \sum_{\substack{a_3\in\mathcal{Q}(1,y)\\ (a_3,b)=1}} \frac{\mu^2(a_3)}{a_3} \sum_{\substack{a_2\in\mathcal{P}(q_0,y)\\ (a_2,b)=1}} \frac{\mu^2(a_2)}{a_2} \\ &\qquad \times \sum_{\substack{a_1\in\mathcal{P}(1,q_0),\,(a_1,b)=1\\ u-f(a_2a_3b)<f(a_1)\le u-f(a_2b)+\epsilon}} \frac{\mu^2(a_1)}{a_1} \\ &\le \sum_{\substack{P^+(b)\le y\\ b\,\text{square-full}}} \frac{1}{b} \sum_{a_3\in\mathcal{Q}(1,y)} \frac{1}{a_3} \sum_{a_2\in\mathcal{P}(q_0,y)} \frac{1}{a_2} \sup_{v\in\mathbb{R}} \Bigg\{ \sum_{\substack{a_1\in\mathcal{P}(1,q_0)\\ v<f(a_1)\le v+\epsilon}} \frac{\mu^2(a_1)}{a_1} \Bigg\} \\ &\ll \frac{\log y}{\log q_0} \sup_{v\in\mathbb{R}} \Bigg\{ \sum_{\substack{a_1\in\mathcal{P}(1,q_0)\\ v<f(a_1)\le v+\epsilon}} \frac{\mu^2(a_1)}{a_1} \Bigg\}. \end{aligned}$$

Next, fix $v \in \mathbb{R}$ and let

$$\mathcal{N} = \{n \in \mathcal{P}(1, q_0) : \mu^2(n) = 1,\ v < f(n) \le v + \epsilon\}.$$

As in the proof of Theorem 1.3, we split $\mathcal{N}$ according to the size of $\min\{p \,|\, n : p > q_J\}$. So we write $\mathcal{N} = \bigcup_{j=0}^{J} \mathcal{N}_j$, where $\mathcal{N}_0 = \mathcal{N} \cap \mathcal{P}(1, q_J)$ and

$$\mathcal{N}_j = \{n \in \mathcal{N} \setminus \mathcal{N}_0 : n = apb,\ P^+(a) \le q_J < p < P^-(b),\ q_j < p \le q_{j-1}\}.$$

First, we bound $\sum_{n\in\mathcal{N}_0} 1/n$. If $n > 1$, then we write $n = mP^+(n) = mp'$. So we find that

$$\tag{4.7} \begin{aligned} \sum_{n\in\mathcal{N}_0} \frac{1}{n} &\le 1 + \sum_{P^+(m)\le q_J} \frac{1}{m} \sum_{\substack{p'\in\mathcal{P}\cap(P^+(m),q_J]\\ v-f(m)<f(p')\le v-f(m)+\epsilon}} \frac{1}{p'} \\ &\ll 1 + \sum_{P^+(m)\le q_J} \frac{1}{m} \left(\frac{\epsilon}{2^J\delta} + \frac{1}{\log^2(1 + P^+(m))} \right) \ll 1 + \frac{\epsilon \log q_J}{2^J\delta}, \end{aligned}$$

by applying the second part of (4.3) with $v - f(m)$, $2^J\delta$, $P^+(m)$ and q_J in place of u, δ, z and w, respectively.

Next, we bound $\sum_{n\in\mathcal{N}_j} 1/n$ for $j \in \{1,\dots,J\}$. In this part of the argument we may assume that $J \geq 1$; otherwise, there is no such j. We have

(4.8)
$$\sum_{n\in\mathcal{N}_j}\frac{1}{n} \leq \sum_{b\in\mathcal{P}(q_j,q_0)}\frac{1}{b}\sum_{a_1\in\mathcal{P}(q_J^\lambda,q_J)}\frac{1}{a_1}\sum_{a_2\in\mathcal{P}(1,q_J^\lambda)}\frac{\mu^2(a_2)}{a_2}\sum_{\substack{p\in\mathcal{P}\cap(q_j,q_{j-1}]\\ v<f(p)+f(a_1b)+f(a_2)\leq v+\epsilon}}\frac{1}{p}$$
$$\leq \sum_{b\in\mathcal{P}(q_j,q_0)}\frac{1}{b}\sum_{a_1\in\mathcal{P}(q_J^\lambda,q_J)}\frac{1}{a_1}\sup_{t\in\mathbb{R}}\Bigg\{\sum_{a_2\in\mathcal{P}(1,q_J^\lambda)}\frac{\mu^2(a_2)}{a_2}\sum_{\substack{p\in\mathcal{P}\cap(q_j,q_{j-1}]\\ t<f(p)+f(a_2)\leq t+\epsilon}}\frac{1}{p}\Bigg\}.$$

Fix some $t \in \mathbb{R}$ and consider $a_2 \in \mathcal{P}(1, q_J^\lambda)$ and $p \in \mathcal{P} \cap (q_j, q_{j-1}]$ with $t < f(a_2) + f(p) \leq t + \epsilon$, as above. Since $|f(p)| \leq 2^j\delta$ for $p \in \mathcal{P} \cap (q_j,\infty)$, by (4.2), we must have $|f(a_2) - t| \leq 2^{j+1}\delta$. So

$$\sum_{a_2\in\mathcal{P}(1,q_J^\lambda)}\frac{\mu^2(a_2)}{a_2}\sum_{\substack{p\in\mathcal{P}\cap(q_j,q_{j-1}]\\ t<f(p)+f(a_2)\leq t+\epsilon}}\frac{1}{p} \tag{4.9}$$
$$= \sum_{\substack{a_2\in\mathcal{P}(1,q_J^\lambda)\\ |f(a_2)-t|\leq 2^{j+1}\delta}}\frac{\mu^2(a_2)}{a_2}\sum_{\substack{p\in\mathcal{P}\cap(q_j,q_{j-1}]\\ t-f(a_2)<f(p)\leq t-f(a_2)+\epsilon}}\frac{1}{p}$$
$$\ll \frac{\epsilon}{2^j\delta}\sum_{\substack{a_2\in\mathcal{P}(1,q_J^\lambda)\\ |f(a_2)-t|\leq 2^{j+1}\delta}}\frac{\mu^2(a_2)}{a_2},$$

by the first part of (4.3) applied with $t - f(a_2)$, $2^{j-1}\delta$, q_j and q_{j-1} in place of u, δ, z and w, respectively, since $z \geq q_j \geq q_J = P_f(2^J\delta) \geq (2^J\delta/\epsilon)^\rho \geq (2^j\delta/\epsilon)^\rho$. Finally, if $a_2 > 1$, then we write $a_2 = mP^+(a_2) = mp'$. Consequently,

$$\sum_{\substack{a_2\in\mathcal{P}(1,q_J^\lambda)\\ |f(a_2)-w|\leq 2^{j+1}\delta}}\frac{\mu^2(a_2)}{a_2} \leq 1 + \sum_{m\in\mathcal{P}(1,q_J^\lambda)}\frac{1}{m}\sum_{\substack{p'\in\mathcal{P}\cap(P^+(m),q_J^\lambda]\\ |f(p')-(t-f(m))|\leq 2^{j+1}\delta}}\frac{1}{p'}.$$

For every fixed $m \in \mathbb{N}$ we have

$$\sum_{\substack{p'\in\mathcal{P}\cap(P^+(m),q_J^\lambda]\\ |f(p')-(t-f(m))|\leq 2^{j+1}\delta}}\frac{1}{p'} \ll \frac{2^j\delta}{2^J\delta} + \frac{1}{\log^2(1+P^+(m))},$$

by (4.3) with $2^j\delta$, $2^J\delta$, $P^+(m)$ and q_J^λ in place of ϵ, δ, z and w, respectively, and with $u\in\{t-f(m)+h\cdot 2^j\delta : h\in\{-2,-1,0,1\}\}$. So we find that

$$\sum_{\substack{a_2\in\mathcal{P}(1,q_J^\lambda)\\ |f(a_2)-w|\le 2^{j+1}\delta}} \frac{\mu^2(a_2)}{a_2} \ll 1+\sum_{P^+(m)\le q_J^\lambda}\frac{1}{m}\left(\frac{2^j\delta}{2^J\delta}+\frac{1}{\log^2(1+P^+(m))}\right)$$

$$\ll 1+\frac{1+\lambda\log q_J}{2^{J-j}}\ll 1+\frac{\lambda\log q_J}{2^{J-j}}.$$

Combining the above estimate with (4.8) and (4.9) implies that

$$\sum_{n\in\mathcal{N}_j}\frac{1}{n}\ll\frac{\epsilon}{2^j\delta}\left(1+\frac{\lambda\log q_J}{2^{J-j}}\right)\sum_{b\in\mathcal{P}(q_j,q_0)}\frac{1}{b}\sum_{a_1\in\mathcal{P}(q_J^\lambda,q_J)}\frac{1}{a_1}$$

$$\ll\frac{\epsilon}{2^j\delta}\left(1+\frac{\lambda\log q_J}{2^{J-j}}\right)\frac{\log q_0}{\log q_j}\,\frac{1}{\lambda}\le\frac{\epsilon}{2^j\delta}\left(\frac{1}{\lambda}+\frac{\log q_J}{2^{J-j}}\right)\frac{\log q_0}{\log q_j},$$

which, together with (4.6) and (4.7), completes the proof of the lemma. ∎

We are now in a position to complete the proof of Theorem 1.5.

Proof of the upper bound in Theorem 1.5. As we have already seen, the function $t\mapsto g(t)(\log t)^c$ is decreasing. In particular, g is strictly decreasing. For every $\delta\in(0,1]$, we define

$$K^*(\delta)=\min\{n\in\mathbb{N}: n\ge 3,\ g(n)\le\delta\}.$$

Then $K^*(\delta)-1\le K(\delta)\le 2K^*(\delta)$, the second inequality being a consequence of Bertrand's postulate.

We claim that

$$\log(K^*(\delta)-1)\ge\frac{1}{2}\left(\frac{\eta}{\delta}\right)^{1/c}\log(K^*(\eta)-1)\qquad(0<\eta\le\delta\le 1). \tag{4.10}$$

Indeed, if $K^*(\delta)=K^*(\eta)$, then this inequality holds trivially. Next, assume that $K^*(\eta)\ge K^*(\delta)+1\ge 4$. Then the definition of $K^*(\eta)$ implies that $g(K^*(\eta)-1)>\eta$. Since, in addition, the function $t\mapsto g(t)(\log t)^c$ is decreasing and $(x-1)\le x^2$ for all $x\ge 3$, we find that

$$1\ge\frac{g(K^*(\eta)-1)(\log(K^*(\eta)-1))^c}{g(K^*(\delta))(\log K^*(\delta))^c}\ge\frac{\eta(\log(K^*(\eta)-1))^c}{\delta(\log K^*(\delta))^c}$$

$$\ge\frac{\eta(\log(K^*(\eta)-1))^c}{\delta(2\log(K^*(\delta)-1))^c}.$$

In any case, (4.10) holds.

Using relation (4.10), we shall show that we may apply Lemma 4.1 with $P_f=K^*-1$, $\mathcal{P}$, $\lambda=1/A$ and $\rho=2$. Condition (4.1) holds by assumption, and condition (4.2) follows immediately by the definition of K^* and the fact that $|f(p)|\le g(K^*(\delta))\le\delta$ for $p\ge K^*(\delta)$. Lastly, we show (4.3) with

$\lambda = 1/A$ and $\rho = 2$. This will be done in several steps. Fix $u \in \mathbb{R}$, $0 < \eta \le \delta \le 1$ and $2 \le z \le w \le \min\{P_f(\delta), P_f(\eta)^{1/A}\}$.

First, we show (4.3) when $z \ge w^{1/4}$. By assumption, there is an absolute constant $C > 0$ such that

$$|f(p_1) - f(p_2)| \ge \frac{2}{C} \min\left\{ \frac{g(p_2)(p_2 - p_1)}{p_2 \log p_2}, g(p_2^A) \right\} \quad (p_1 < p_2,\ p_1, p_2 \in \mathbb{P}).$$

We claim that if $v \in \mathbb{R}$ and $\eta' := \min\{\eta, \delta/\log w\}/C$, then

$$\sum_{\substack{p \in \mathcal{P} \cap (z,w] \\ v < f(p) \le v + \eta'}} \frac{1}{p} \ll \begin{cases} \eta'/\delta & \text{if } z \ge P_f(2\delta) \ge (2\delta/\eta)^2, \\ \eta'/\delta + 1/(\log w)^3 & \text{otherwise.} \end{cases} \tag{4.11}$$

If this relation does hold, then breaking the interval $(u, u+\eta]$ into at most $1 + \eta/\eta' \le 1 + C \log w$ intervals of the form $(v, v+\eta']$, we deduce that (4.3) holds too when $z \ge w^{1/4}$. So it remains to show (4.11) to complete the proof of (4.3) in this special case.

Without loss of generality, we may assume that $w \ge 3$; otherwise there are no primes in $(z, w] \subset (2, 3)$ and (4.11) is trivially true. In particular, we may assume that $K^*(\eta) \ge K^*(\delta) \ge 4$. Therefore, for every $p \in (z, w]$, we see that $g(p) \ge g(K^*(\eta) - 1) > \eta$ and $g(p^A) \ge g(K^*(\delta) - 1) > \delta$. Now, consider two primes $p_1 < p_2$ that both belong to the set $\{p \in \mathcal{P} \cap (z, w] : v < f(p) \le v + \eta'\}$. Then

$$\begin{aligned} \frac{1}{C} \min\left\{\eta, \frac{\delta}{\log w}\right\} = \eta' &> |f(p_1) - f(p_2)| \\ &\ge \frac{2}{C} \min\left\{ \frac{g(p_2)(p_2 - p_1)}{p_2 \log p_2}, g(p_2^A) \right\} \\ &\ge \frac{2}{C} \min\left\{ \frac{\delta(p_2 - p_1)}{p_2 \log p_2}, \eta \right\}, \end{aligned} \tag{4.12}$$

and consequently

$$0 < p_2 - p_1 \le \frac{C\eta'}{2\delta} p_2 \log p_2 = \min\left\{\frac{\eta}{\delta}, \frac{1}{\log w}\right\} \frac{p_2 \log p_2}{2} \le \frac{p_2}{2}.$$

Set

$$P = \max\{p \in \mathcal{P} \cap (z, w] : v < f(p) \le v + \eta'\} \quad \text{and}$$
$$y = \frac{C\eta'}{2\delta} P \log P \le \frac{P}{2},$$

so that $\{p \in \mathcal{P} \cap (z, w] : v < f(p) \le v + \eta'\} \subset [P - y, P]$. The second part of relation (4.11) then follows by the Prime Number Theorem [13, Theorem 1, p. 167]. For the first part of (4.11), note that if $z \ge P_f(2\delta) \ge (2\delta/\eta)^2$, then

$$\frac{y}{\sqrt{P}} = \frac{\sqrt{P} \log P}{2} \min\left\{\frac{\eta}{\delta}, \frac{1}{\log w}\right\} \ge \frac{(2\delta/\eta) \log z}{2} \min\left\{\frac{\eta}{\delta}, \frac{1}{\log w}\right\} \ge \frac{1}{4},$$

where we used our assumption that $z \geq \max\{w^{1/4}, 2\}$. So the first part of (4.11) follows by the Brun–Titchmarsch inequality [13, Theorem 9, p. 73], thus completing the proof of (4.11) and hence of (4.3) in the case when $z \geq w^{1/4}$.

Finally, we show (4.3) when $z < w^{1/4}$. First, note that

$$\log P_f(2^j\delta) \geq 2^{-1-j/c} \log P_f(\delta) \geq 2^{-1-j/c} \log w \geq 4^{-j} \log w,$$

by (4.10), for every $j \geq 1$. Since $w \leq P_f(\delta)$ too, by assumption, we deduce that

$$P_f(2^j\delta) \geq w_j := w^{4^{-j}} \qquad (j \geq 0). \tag{4.13}$$

Applying this inequality with $j = 1$ implies that $z < P_f(2\delta)$, that is to say, we are in the second case of (4.3). Let

$$\begin{aligned} j_0 &= \max\{j \geq 0 : w_j \geq z \text{ and } 2^j \leq 1/\delta\}, \\ S_j &= \sum_{\substack{p \in \mathcal{P} \cap (w_{j+1}, w_j] \\ u < f(p) \leq u+\eta}} \frac{1}{p} \quad \text{ for } j \in \{0, 1, \ldots, j_0 - 1\} \text{ and} \\ S_{j_0} &= \sum_{\substack{p \in \mathcal{P} \cap (z, w_{j_0}] \\ u < f(p) \leq u+\eta}} \frac{1}{p}. \end{aligned}$$

Then the part of (4.3) that we have already proven and (4.13) imply that

$$S_j \ll \frac{\eta}{2^j\delta} + \frac{16^j}{(\log w)^2}, \tag{4.14}$$

for $j \in \{0, 1, \ldots, j_0 - 1\}$. We claim that the same estimate holds for S_{j_0}. If $2^{j_0+1}\delta \leq 1$, then $w_{j_0}^{1/4} = w_{j_0+1} < z$ and thus we may apply again the part of (4.3) that we have already proven. Finally, if $2^{j_0+1}\delta > 1$, then we have $w_{j_0} \leq P_f(2^{j_0}\delta) \leq P_f(1/2) \ll 1$, since $g(t) \ll 1/(\log t)^c$ by our assumptions on f. Consequently, covering the interval $(z, w_{j_0}]$ by $O(1)$ intervals of the form $(t, t^4]$ and applying the already proven part of (4.3) shows that (4.14) holds in this case too for $j = j_0$. Summing (4.14) over $j \in \{0, 1, \ldots, j_0\}$ implies that

$$\sum_{\substack{p \in \mathcal{P} \cap (z, w] \\ u < f(p) \leq u+\eta}} \frac{1}{p} = \sum_{j=0}^{j_0} S_j \ll \frac{\eta}{\delta} + \frac{1}{(\log z)^2},$$

which completes the proof of (4.3). In conclusion, we may apply Lemma 4.1 with $P_f = K^* - 1$, $\rho = 2$ and $\lambda = 1/A$.

We are finally ready to show the upper bound in Theorem 1.5. Let $\epsilon \in [0, 1/2]$. We may assume that $K^*(\epsilon)$ is large enough; otherwise, the theorem follows by the trivial upper bound $Q_{\mathcal{F}_y}(\epsilon) \leq 1$. In particular, we may assume that the parameter $\delta := g(\lfloor K^*(\epsilon)^{1/A}\rfloor - 1)$ lies in $[\epsilon, 1/2]$. Since g is strictly decreasing, the definition of K^* implies that

$$K^*(\delta) = \lfloor K^*(\epsilon)^{1/A}\rfloor - 1. \tag{4.15}$$

In particular, $P_f(\delta) \le P_f(\epsilon)^{1/A}$. For $j \in \mathbb{N}\cup\{0\}$ with $2^j \le 1/\delta$, we set $q_j = P_f(2^j\delta) = K^*(2^j\delta) - 1$. Note that

$$q_0 = K^*(\delta) - 1 \le K^*(\epsilon) - 1 \le K(\epsilon)$$

and

$$\log q_j \ge 2^{-1-(j-i)/c}\log q_i \qquad (0 \le i \le j \le \log(1/\delta)/\log 2), \tag{4.16}$$

by (4.10). Set

$$J = \max\{\{0\}\cup\{j\in\mathbb{N} : 2^j \le 1/\delta \text{ and } q_j \ge (2^j\delta/\epsilon)^2\}\}.$$

Then Lemma 4.1 and (4.16) imply that, for $y \ge K(\epsilon) \ge q_0$,

$$\begin{aligned} Q_{\mathcal{F}_y}(\epsilon) &\ll_A \frac{1}{\log q_0} + \frac{\epsilon}{\delta}\sum_{j=1}^{J}\frac{1}{2^j\log q_j} + \frac{\epsilon}{\delta}\sum_{j=0}^{J}\frac{\log q_J}{2^J\log q_j} \\ &\ll \frac{1}{\log q_0} + \frac{\epsilon}{\delta}\sum_{j=1}^{J}\frac{1}{2^{j-j/c}\log q_0} + \frac{\epsilon}{\delta}\sum_{j=0}^{J}\frac{\log q_J}{2^{J-j/c}\log q_0} \\ &\ll \frac{\min\{1/(c-1), 1+J\epsilon/\delta\}}{\log q_0} + \frac{\epsilon}{\delta}\,\frac{\log q_J}{2^{J(1-1/c)}\log q_0} \\ &\ll \frac{\min\{1/(c-1), 1+J\epsilon/\delta\}}{\log q_0} + \frac{\epsilon}{\delta}\,\frac{\log q_J}{\max\{1,(J+1)(c-1)\}\log q_0}. \end{aligned} \tag{4.17}$$

Finally, note that if $2^{J+1} \le 1/\delta$, then the maximality of J and (4.16) imply that

$$\log q_J \le 4\log q_{J+1} \le 8\log(2^{J+1}\delta/\epsilon) \ll J + 1 + \log(\delta/\epsilon).$$

On the other hand, if $2^{J+1} > 1/\delta$, then $q_J \le K^*(1/2) \ll 1$, since $g(t) \ll (\log t)^{-c}$. In any case, we find that $\log q_J \ll J+1+\log(\delta/\epsilon)$. So the inequalities

$$\frac{J\epsilon}{\delta} \ll \frac{\epsilon\log(1/\delta)}{\delta} \le \log(1/\epsilon) \quad \text{and} \quad \frac{\epsilon}{\delta}\log(\delta/\epsilon) \ll 1$$

and relation (4.17) imply that

$$Q_{\mathcal{F}_y}(\epsilon) \ll \min\{1/(c-1), \log(1/\epsilon)\}/\log q_0.$$

Finally, $\log q_0 \asymp_A \log K^*(\epsilon) \asymp \log K(\epsilon)$, by (4.15) and the fact that $K^* - 1 \le K \le 2K^*$. So the upper bound in Theorem 1.5 follows. ∎

Acknowledgments. I would like to thank Gérald Tenenbaum for pointing out relation (1.4). I am also grateful to Régis de la Bretèche and Maksym Radziwill for some helpful comments. In addition, I would like to thank the referees who handled the paper, as their comments exposed some inaccuracies and improved the exposition of the main ideas. This paper was largely written while I was visiting Université Paris-Diderot, which I would like to thank for its hospitality. During that time I held a postdoctoral fellowship at the Centre de recherches mathématiques at Montréal, which I would like to thank for the financial support.

References

[1] R. de la Bretèche and G. Tenenbaum, *Sur la concentration de certaines fonctions additives*, Math. Proc. Cambridge Philos. Soc. 152 (2012), 179–189; Corrigendum, ibid., 191.

[2] H. Diamond and D. Rhoads, *The modulus of continuity of the distribution function of $\phi(n)/n$*, in: Topics in Classical Number Theory, Vol. I (Budapest, 1981), Colloq. Math. Soc. János Bolyai 34, North-Holland, Amsterdam, 1984, 335–353.

[3] P. D. T. A. Elliott, *Probabilistic Number Theory. I. Mean-Value Theorems*, Grundlehren Math. Wiss. 239, Springer, Berlin, 1979.

[4] P. D. T. A. Elliott, *Probabilistic Number Theory. II. Central Limit Theorems*, Grundlehren Math. Wiss. 240, Springer, Berlin, 1980.

[5] P. Erdős, *On the distribution function of additive functions*, Ann. of Math. (2) 47 (1946), 1–20.

[6] P. Erdős, *On the distribution of numbers of the form $\sigma(n)/n$ and on some related questions*, Pacific J. Math. 52 (1974), 59–65.

[7] P. Erdős and I. Kátai, *On the concentration of distribution of additive functions*, Acta Sci. Math. (Szeged) 41 (1979), 295–305.

[8] P. Erdős and A. Wintner, *Additive arithmetical functions and statistical independence*, Amer. J. Math. 61 (1939), 713–721.

[9] G. Halász, *On the distribution of additive arithmetic functions*, in: Collection of Articles in Memory of Juriĭ Vladimirovič Linnik, Acta Arith. 27 (1975), 143–152.

[10] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Math. 90, Cambridge Univ. Press, Cambridge, 1988.

[11] J. Kubilius, *Probabilistic Methods in the Theory of Numbers*, Transl. Math. Monogr. 11, Amer. Math. Soc., Providence, RI, 1964.

[12] I. Z. Ruzsa, *On the concentration of additive functions*, Acta Math. Acad. Sci. Hungar. 36 (1980), 215–232.

[13] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Stud. Adv. Math. 46, Cambridge Univ. Press, Cambridge, 1995.

[14] M. M. Tjan, *On the question of the distribution of values of the Euler function $\varphi(n)$*, Litovsk. Mat. Sb. 6 (1966), 105–119 (in Russian).

Dimitris Koukoulopoulos
Département de Mathématiques et de Statistique
Université de Montréal
CP 6128 succ. Centre-Ville
Montréal, Québec H3C 3J7, Canada
E-mail: koukoulo@dms.umontreal.ca

Received on 24.2.2012
and in revised form on 9.11.2013 (6986)