Journal of Integer Sequences, Vol. 16 (2013),
Article 13.5.5

On the Middle Prime Factor of an Integer

Jean-Marie De Koninck
Département de Mathématiques
Université Laval
Québec G1V 0A6
Canada
jmdk@mat.ulaval.ca

Florian Luca
Fundación Marcos Moshinsky
UNAM
Circuito Exterior C.U.
Apdo. Postal 70-543
Mexico D.F. 04510
Mexico
fluca@matmor.unam.mx

Abstract

Given an integer $n \geq 2$, let $p_m(n)$ denote the middle prime factor of n. We obtain an estimate for the sum of the reciprocals of $p_m(n)$ for $n \leq x$.

1 Introduction

Given an integer $n \geq 2$, let $P(n)$ stand for its largest prime factor, writing $P(1) = 1$. The global behavior of this function is easy to evaluate. For instance, one can easily show that

$$\sum_{n \leq x} P(n) = \frac{\pi^2}{12} \frac{x^2}{\log x} + O\left(\frac{x^2}{\log^2 x}\right).$$

This is a well known result and a proof can be found in our recent book ([4, Thm. 9.2]).

Estimating the global behavior of the reciprocal of $P(n)$ is much harder, and it has been the focus of many papers at the end of the 1970's and early 1980's. See, for instance, the book of De Koninck and Ivić [2] and the papers of Erdős and Ivić [5, 6]. The best estimate was finally obtained in 1986 by Erdős, Ivić and Pomerance [7]; they proved that

$$\sum_{n\le x}\frac{1}{P(n)} = x\delta(x)\left(1+O\left(\sqrt{\frac{\log_2 x}{\log x}}\right)\right),$$

where $\delta(x)$ is some continuous function which decreases to 0 very slowly as $x\to\infty$ and satisfies

$$\delta(x)=\exp\{-\sqrt{2\log x\log_2 x}(1+o(1))\}\qquad \text{as } x\to\infty.$$

Here and in what follows, for an integer $k\ge 2$, we write $\log_k x$ for the k-th fold iterate of log evaluated at x, and we shall assume that the input x in such an expression is sufficiently large such that the thus defined iterated logarithm is real and positive.

Interestingly the sum of the reciprocals of the second largest prime factor function $P_2(n)$ defined for those integers n with at least two prime factors, has a totally different asymptotic value. In fact, the first author proved in 1993 (see De Koninck [1]) that, if $\Omega(n)$ stands for the number of prime factors of n counting their multiplicity,

$$\sum_{\substack{n\le x\\ \Omega(n)\ge 2}}\frac{1}{P_2(n)} = c_2\frac{x}{\log x}+O\left(\frac{x}{\log^2 x}\right),$$

where $c_2=\displaystyle\sum_{m=1}^{\infty}\frac{1}{m}\sum_{p\ge P(m)}\frac{1}{p^2}\approx 1.254$. In the same paper, it is shown that, if $P_k(n)$ stands for the k-largest prime factor of an integer, then

$$\sum_{\substack{n\le x\\ \Omega(n)\ge k}}\frac{1}{P_k(n)} = c_k\frac{x(\log_2 x)^{k-2}}{\log x}\left(1+O\left(\frac{1}{\log_2 x}\right)\right),$$

where $c_k=c_2/(k-2)!$.

In 1988, De Koninck and Galambos [3] estimated the sum of the reciprocals of a random prime factor of an integer. More precisely, for each integer $n\ge 2$, let

$$p_1(n)<p_2(n)<\cdots<p_{\omega(n)}(n)$$

be its $k=\omega(n)$ distinct prime factors, and then select at random (with equal probability $1/k$) one of these prime factors and call it $p_*(n)$. Then, set

$$R(x)=\sum_{2\le n\le x}\frac{1}{p_*(n)}. \qquad (1)$$

Note that the total number of sums of the above form is $\omega(2)\omega(3)\cdots\omega(\lfloor x\rfloor)$. We shall say that a property holds for almost all sums in (1) if the number $N(x)$ of sums with the property in question satisfies

$$\frac{N(x)}{\omega(2)\omega(3)\cdots\omega(\lfloor x\rfloor)} \to 1 \qquad \text{as } x \to \infty.$$

Then they proved that, for almost all sums in (1),

$$R(x) = c_1 \frac{x}{\log_2 x} + O\left(\frac{x}{(\log_2 x)^2}\right),$$

where $c_1 = \sum_p \frac{1}{p^2} \approx 0.452$.

During the 1984 Oberwolfach Conference on Analytic Number Theory, Paul Erdős asked the first author if he had thought of estimating the sum of the reciprocals of the middle prime factor of an integer. After almost 30 years, we can now prove an estimate for this sum.

2 The main result

Given an integer $n \geq 2$, write it as $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$, where $p_1 < p_2 < \cdots < p_k$ are its distinct prime factors and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive integers. We let $p_m(n) = p_{\max(1,\lfloor k/2\rfloor)}$ and say that $p_m(n)$ is the "middle" prime factor of n. We will prove the following estimate.

Theorem 1. *As $x \to \infty$,*

$$\sum_{n \leq x} \frac{1}{p_m(n)} = \frac{x}{\log x} \exp\left((c_0 + o(1))\sqrt{\log_2 x \, \log_3 x}\right),$$

where $c_0 = \sqrt{2}$.

3 The proof of the upper bound

For the upper bound, we first set

$$\begin{aligned} \mathcal{N}_1(x) &= \{n \leq x : \Omega(n) > 10 \log_2 x\}; \\ \mathcal{N}_2(x) &= \{n \leq x : p_m(n) > \log x\}; \\ \mathcal{N}_3(x) &= \{n \leq x : \omega(n) = 1,\ 2\}. \end{aligned}$$

By Luca and Pomerance [9, Lem. 13], we have

$$\#\mathcal{N}_1(x) \ll \frac{x \log x \log_2 x}{2^{10 \log_2 x}} \ll \frac{x}{(\log x)^5},$$

which shows that

$$\sum_{n\in\mathcal{N}_1(x)} \frac{1}{p_m(n)} \ll \#\mathcal{N}_1(x) \ll \frac{x}{(\log x)^5}. \tag{2}$$

Clearly,

$$\sum_{n\in\mathcal{N}_2(x)} \frac{1}{p_m(n)} \le \frac{\#\mathcal{N}_2(x)}{\log x} \ll \frac{x}{\log x}. \tag{3}$$

Finally,

$$\#\mathcal{N}_3(x) \ll \frac{x\log_2 x}{\log x},$$

so that

$$\sum_{n\in\mathcal{N}_3(x)} \frac{1}{p_m(n)} \ll \frac{x\log_2 x}{\log x}. \tag{4}$$

In light of these estimates for the contributions of the members n from $\mathcal{N}_1(x)$, $\mathcal{N}_2(x)$ and $\mathcal{N}_3(x)$ to the sum of the reciprocals of $p_m(n)$, from now on we assume that $n \in \mathcal{N}_4(x) = \{n \le x\}\backslash\left(\mathcal{N}_1(x)\cup\mathcal{N}_2(x)\cup\mathcal{N}_3(x)\right)$.

Fix $p \in [2, \log x]$ and $k \in [3, 10\log_2 x]$ and consider the set

$$\mathcal{N}_{p,k}(x) = \{n \in \mathcal{N}_4(x) : p_m(n) = p \text{ and } \omega(n) = k\}.$$

Write $k = 2k_0 + \delta$, where $\delta \in \{0, 1\}$. If $n \in \mathcal{N}_{p,k}(x)$, then $n = ap^\alpha b$, where $P(a) < p$, $\alpha \ge 1$, all prime factors of b exceed p and $\omega(a) = k_0 - 1$, and $\omega(b) = k_0 + \delta$. Fix also a. Then $b \le x/ap^\alpha$. Note that

$$ap^\alpha \le p^{10\log_2 x} < (\log x)^{10\log_2 x} = x^{o(1)} \quad \text{as} \quad x \to \infty.$$

Hence, given a and p^α, the number of choices for the number b is, by the Hardy-Ramanujan inequalities (see Hardy and Ramanujan [8] or De Koninck and Luca [4, Thm. 10.1]),

$$\begin{aligned} &\ll \frac{x}{ap^\alpha\log(x/ap^\alpha)} \times \frac{1}{(k_0+\delta-1)!} \times \left(\log_2(x/ap^\alpha) + O(1)\right)^{k_0+\delta-1} \\ &\ll \frac{x(\log_2 x)}{ap^\alpha\log x} \times \frac{1}{(k_0-1)!} \times (\log_2 x)^{k_0-1}. \end{aligned}$$

Thus, since $k! \geq (k/e)^k$, we have

$$\begin{aligned}
\frac{1}{p}\#\mathcal{N}_{p,k}(x) &\leq \frac{1}{p} \sum_{\substack{a \\ \omega(a)=k_0-1 \\ \Omega(a)<10\log_2 x \\ P(a)<p}} \sum_{\alpha\geq 1} \frac{x(\log_2 x)}{ap^{\alpha}\log x} \times \frac{1}{(k_0-1)!} \times (\log_2 x)^{k_0-1} \\
&\ll \frac{x(\log_2 x)}{p^2\log x} \times \frac{(\log_2 x)^{k_0-1}}{(k_0-1)!} \sum_{\substack{a \\ \omega(a)=k_0-1 \\ \Omega(a)<10\log_2 x \\ P(a)<p}} \frac{1}{a} \\
&\ll \frac{x(\log_2 x)}{p^2\log x} \times \frac{(\log_2 x)^{k_0-1}}{(k_0-1)!} \times \frac{1}{(k_0-1)!} \left(\sum_{\substack{q<p \\ \beta\geq 1}} \frac{1}{q^{\beta}} \right)^{k_0-1} \\
&\ll \frac{x(\log_2 x)}{p^2\log x} \times \frac{(\log_2 x)^{k_0-1}}{(k_0-1)!} \times \frac{(\log_2 p)^{k_0-1}}{(k_0-1)!} \\
&\ll \frac{x(\log_2 x)^2}{p^2\log x} \left(\frac{e^2\log_2 x\log_2 p}{k_0^2} \right)^{k_0}.
\end{aligned}$$

For fixed $A>1$, the maximum of the function $f_A(t) = (e^2A/t^2)^t$ is attained at $t_0 = \sqrt{A}$ in which case $f_A(t_0) = \exp(2\sqrt{A})$. We first take $B = \exp(2\sqrt{\log_2 x \log_3 x})$ and split the range of p into values $p \leq B$ and $p > B$. Let $\mathcal{N}_5(n)$ and $\mathcal{N}_6(n)$ be the subsets of $\mathcal{N}_4(x)$ for which $p(n) \leq B$ and $p(n) > B$, respectively. Applying the above observation with

$$A = \log_2 x \, \log_2 B,$$

we get that

$$\begin{aligned}
\sum_{n\in\mathcal{N}_5(x)} \frac{1}{p_m(n)} &\leq \sum_{\substack{p\leq B \\ 3\leq k<10\log_2 x}} \frac{1}{p}\#\mathcal{N}_{p,k}(x) \\
&\ll \sum_{\substack{p\leq B \\ 1\leq k_0<5\log_2 x}} \frac{x(\log_2 x)^2}{p^2\log x} \left(\frac{e^2\log_2 x\log_2 B}{k_0^2} \right)^{k_0} \\
&\ll \frac{x(\log_2 x)^3}{\log x} \exp(2\sqrt{A}) \left(\sum_{2\leq p\leq B} \frac{1}{p^2} \right) \\
&\ll \frac{x}{\log x} \exp\left((c_0+o(1))\sqrt{\log_2 x\log_3 x} \right) \qquad (5)
\end{aligned}$$

as $x \to \infty$. For $\mathcal{N}_6(x)$, we use the fact that $p \leq \log x$, put $C = \log_2 x \log_3 x$, and use the above argument based on the maximum of the function $f_C(t)$ to get that

$$\begin{aligned}\sum_{n\in\mathcal{N}_6(x)}\frac{1}{p_m(n)} &\ll \sum_{\substack{B<p<\log x\\ 3\le k<10\log_2 x}}\frac{1}{p}\#\mathcal{N}_{p,k}(x)\\ &\ll \sum_{\substack{B<p\le\log x\\ 1\le k_0<5\log_2 x}}\frac{x(\log_2 x)^2}{p^2\log x}\left(\frac{e^2\log_2 x\log_3 x}{k_0^2}\right)^{k_0}\\ &\ll \frac{x(\log_2 x)^3}{\log x}\exp(2\sqrt{C})\left(\sum_{p>B}\frac{1}{p^2}\right)\ll\frac{x(\log_2 x)^3}{\log x},\end{aligned}\tag{6}$$

because $B=\exp(2\sqrt{C})$. Comparing estimates (2), (3), (4), (5) and (6), we get the desired estimate.

4 The proof of the lower bound

We now turn our attention to the lower bound. For this, we shall select a subset of positive integers $n\le x$ such that the sum of the reciprocals of $p_m(n)$ for n's in this subset already achieves the desired lower bound. Our n's will be of the form

$$n=abpP$$

such that $p\in[q_0,2q_0]$, a and b are square free and have k_0 and k_0-1 prime factors, respectively, with $r<p(a)<P(a)\le q_0$ and $2q_0<p(b)<P(b)<x^{1/(2k_0)}$, and P is a prime in the interval $[x/(2pab),x/(pab)]$. Here,

$$q_0=\exp\left(\sqrt{\frac{\log_2 x}{\log_3 x}}\right)\qquad\text{and}\qquad k_0=\lfloor c_0^{-1}\sqrt{\log_2 x\log_3 x}\rfloor,\qquad r=\log_2 x.$$

Formally,

$$\mathcal{A}=\{a:\mu^2(a)=1,\ P(a)\le q_0,\ p(a)>r,\ \omega(a)=k_0\}.$$

We need to estimate the sum of reciprocals of the members of $\mathcal{A}$. For this, we set

$$S=\sum_{r<q\le q_0}\frac{1}{q}\qquad\text{and}\qquad S_1=\sum_{r<q\le q_0}\sum_{\beta\ge1}\frac{1}{q^\beta}.$$

Clearly,

$$S=\log_2 q_0-\log_2 r+O(1)=\frac{1}{2}\log_3 x+O(\log_4 x),\tag{7}$$

and

$$S_1=S+\sum_{r<q\le q_0}\sum_{\beta\ge2}\frac{1}{q^\beta}=S+O\left(\sum_{q>r}\frac{1}{q^2}\right)=S+O\left(\frac{1}{r}\right).$$

In particular,

$$S_1^{k_0} = \left(S + O\left(\frac{1}{r}\right)\right)^{k_0} = S^{k_0}\exp\left(O\left(\frac{k_0}{rS}\right)\right) = S^{k_0}(1+o(1)).$$

Now it is easy to see that

$$\begin{aligned}\sum_{a\in\mathcal{A}}\frac{1}{a} &\geq \frac{1}{k_0!}S^{k_0} - \sum_{r<q<p}\frac{1}{q^2}\frac{S_1^{k_0-1}}{(k_0-1)!}\\ &\geq \frac{S^{k_0}}{k_0!}\left(1+O\left(\frac{k_0}{S}\sum_{q>r}\frac{1}{q^2}\right)\right)\\ &= \frac{S^{k_0}}{k_0!}\left(1+O\left(\frac{k_0}{rS}\right)\right) = \frac{S^{k_0}}{k_0!}(1+o(1)). \qquad (8)\end{aligned}$$

Observe that if $a \in \mathcal{A}$, then $ap < (2q_0)^{k_0+1} < (\log x)^{\log_2 x}$ for all sufficiently large x. Now let

$$\mathcal{B} = \left\{b : \mu^2(b) = 1, p(b) > 2q_0, P(b) < x^{1/(2k_0)}, \omega(b) = k_0 - 1\right\}.$$

Proceeding as in the estimation of the sum of the reciprocals of $a \in \mathcal{A}$, one gets that

$$\sum_{b\in\mathcal{B}}\frac{1}{b} \gg \frac{1}{(k_0-1)!}\left(T + O(1)\right)^{k_0-1} \gg \frac{1}{\sqrt{\log_2 x}}\frac{T^{k_0}}{k_0!}, \qquad (9)$$

where

$$\begin{aligned}T &= \sum_{2q_0<q<x^{1/2k_0}}\frac{1}{q} = \log_2(x^{1/2k_0}) - \log_2(2q_0) + o(1)\\ &= \log_2 x + O(\log_3 x). \qquad (10)\end{aligned}$$

Consider now $a \in \mathcal{A},\ b \in \mathcal{B}$. Then

$$apb < (\log x)^{\log_2 x}x^{(k_0-1)/(2k_0)} < x^{1/2}(\log x)^{\log_2 x} < x^{2/3}$$

for all sufficiently large x, so that letting $P \in [x/(2apb), x/(apb)]$, it follows that

$$P \geq x/(2apb) > x^{1/3} > x^{1/(2k_0)}$$

for large x, implying that given $n = Papb$, the numbers $P,\ a,\ b$ are uniquely determined. Observe that the number of choices for P is

$$\pi\left(\frac{x}{abp}\right) - \pi\left(\frac{x}{2abp}\right) \gg \frac{x}{apb\log(x/apb)} \gg \frac{x}{apb\log x}. \qquad (11)$$

Thus, in light of (8), (9) and (11), it follows that the number of positive integers $n \le x$ such constructed and for which $p_m(n) = p$ is

$$\begin{aligned} &\ge \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \left(\pi\left(\frac{x}{apb}\right) - \pi\left(\frac{x}{2apb}\right)\right) \\ &\gg \frac{x}{p\log x} \left(\sum_{a\in\mathcal{A}} \frac{1}{a}\right) \left(\sum_{b\in\mathcal{B}} \frac{1}{b}\right), \end{aligned}$$

so that

$$\frac{1}{p} \sum_{\substack{n\le x \\ p_m(n)=p}} 1 \gg \frac{x}{p^2 \log x \sqrt{\log\log x}} \frac{(ST)^{k_0}}{k_0!^2},$$

which means, in light of (7) and (10), that

$$\begin{aligned} \frac{1}{p} \sum_{\substack{n\le x \\ p_m(n)=p}} 1 \quad &\gg \quad \frac{x}{p^2(\log x)(\log_2 x)^{3/2}} \\ &\qquad \times \left(\frac{(e^2/2 + o(1))\log_2 x \log_3 x}{(1/2 + o(1))\log_2 x \log_3 x}\right)^{(c_0^{-1}+o(1))\sqrt{\log_2 x \log_3 x}} \\ &\gg \quad \frac{x \exp((c_0 + o(1))\sqrt{\log_2 x \log_3 x})}{\log x} \times \frac{1}{p^2} \end{aligned}$$

implying that

$$\begin{aligned} \sum_{n\le x} \frac{1}{p_m(n)} \quad &\gg \quad \frac{x}{\log x} \exp((c_0 + o(1))\sqrt{\log_2 x \log_3 x}) \sum_{q_0<p<2q_0} \frac{1}{p^2} \\ &\gg \quad \frac{x \exp((c_0 + o(1))\sqrt{\log_2 x \log_3 x})}{\log x} \left(\frac{\pi(2q_0) - \pi(q_0)}{4q_0^2}\right) \\ &\gg \quad \frac{x}{q_0(\log q_0)(\log x)} \exp((c_0 + o(1))\sqrt{\log_2 x \log_3 x}) \\ &\gg \quad \frac{x}{\log x} \exp((c_0 + o(1))\sqrt{\log_2 x \log_3 x}) \qquad (x\to\infty), \end{aligned}$$

as requested.

Remark 2. It would be interesting to obtain bounds on the error terms. One may also ask about variations of this problem in which one considers the median prime, or the middle prime taking into account multiplicities, and so on. We leave all such problems for a future project.

5 Acknowledgments

We thank the referee for suggestions which improved the quality of the manuscript. The work of the first author was supported by a grant from NSERC. The second worked on this paper during a visit to Université Laval in the Fall of 2012 and was supported in part by Projects PAPIIT IN104512, CONACyT Mexico-France 193539, CONACyT Mexico-India 163787, and a Marcos Moshinsky Fellowship.

References

[1] J. M. De Koninck, Sur les plus grands facteurs premiers d'un entier, *Monatsh. Math.* **116** (1993), 13–37.

[2] J. M. De Koninck and A. Ivić, *Topics in Arithmetical Functions*, Mathematics Studies **43**, North-Holland, Amsterdam, 1980.

[3] J. M. De Koninck and J. Galambos, Some randomly selected arithmetical sums, *Acta Math. Hung.* **52** (1988), 37–43.

[4] J. M. De Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers*, Graduate Studies in Mathematics, Vol. 134, American Mathematical Society, 2012.

[5] P. Erdős and A. Ivić, Estimates for sums involving the largest prime factor of an integer and certain related additive functions, *Studia Sci. Math. Hungar.* **15** (1980), 183–199.

[6] P. Erdős and A. Ivić, On sums involving reciprocals of certain arithmetical functions, *Publ. Inst. Math.* (Beograd) **32** (1982), 49–56.

[7] P. Erdős, A. Ivić, and C. Pomerance, On sums involving reciprocals of the largest prime factor of an integer, *Glas. Mat. Ser. III* **21** (1986), 283–300.

[8] G. H. Hardy and S. Ramanujan, The normal number of prime factors of a number n, *Quart. J. Math.* **48** (1917), 76–92.

[9] F. Luca and C. Pomerance, Irreducible radical extensions and Euler-function chains, *Integers* **7** (2007), Paper A25.

2010 *Mathematics Subject Classification*: Primary 11N37.
Keywords: largest prime factor, middle prime factor.

Received March 2 2013; revised version received May 13 2013. Published in *Journal of Integer Sequences*, May 16 2013.

Return to Journal of Integer Sequences home page.