DECOMPOSING A FACTORIAL INTO LARGE FACTORS

BORIS ALEXEEV, ..., TERENCE TAO, AND KEVIN VENTULLO

ABSTRACT. Let $t(N)$ denote the largest number such that $N!$ can be expressed as the product of N numbers greater than or equal to $t(N)$. The bound $t(N)/N = 1/e - o(1)$ was apparently established in unpublished work of Erdős, Selfridge, and Straus; but the proof is lost. Here we obtain the more precise asymptotic

$$\frac{t(N)}{N} = \frac{1}{e} - \frac{c_0}{\log N} + O\left(\frac{1}{\log^{1+c} N}\right)$$

for an explicit constant $c_0 = 0.3044190\dots$ and some absolute constant $c > 0$, answering a question of Erdős and Graham. With numerical assistance, we also establish several conjectures of Guy and Selfridge concerning effective estimates of this quantity, for instance establishing $t(N) \geq N/3$ for $N \geq 43632$, with the threshold shown to be best possible. **This abstract optimistically assumes that the last remaining cases of this conjecture can be resolved.**

1. INTRODUCTION

Given a natural number M, define a *factorization* of M to be a finite multiset $\mathcal{B}$ such that the product

$$\prod \mathcal{B} := \prod_{a \in \mathcal{B}} a$$

(where the elements are counted with multiplicity) is equal to M; more generally, define a *subfactorization* of M to be a finite multiset $\mathcal{B}$ such that $\prod \mathcal{B}$ divides M. Given a threshold t, we say that a multiset $\mathcal{B}$ is *t-admissible* if $a \geq t$ for all $a \in \mathcal{B}$. For a given natural number N, we then define $t(N)$ to be the largest t for which there exists a t-admissible factorization $\mathcal{B}$ of $N!$ of cardinality $|\mathcal{B}| = N$.

Example 1.1. The multiset

$$\{3, 3, 3, 3, 4, 4, 5, 7, 8\}$$

is a 3-admissible factorization of

$$\prod\{3, 3, 3, 3, 4, 4, 5, 7, 8\} = 3^4 \times 4^2 \times 5 \times 7 \times 8 = 9!$$

of cardinality

$$|\{3, 3, 3, 3, 4, 4, 5, 7, 8\}| = 9,$$

hence $t(9) \geq 3$. One can check that no 4-admissible factorization of 9! of this cardinality exists, hence $t(9) = 3$.

2020 *Mathematics Subject Classification.* 11A51.

It is easy to see that $t(N)$ is non-decreasing in N, (any cardinality N factorization of $N!$ can be extended to a cardinality $N+1$ factorization of $(N+1)!$ by adding $N+1$ to the multiset). The first few elements of the sequence $t(N)$ are

$$1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, \dots$$

(OEIS A034258). The values of $t(N)$ for $N \leq 79$ were computed in [9], and the values for $N \leq 200$ can be extracted from OEIS A034259, which describes the inverse sequence to t.

When the factorial $N!$ is replaced with an arbitrary number, this problem is essentially the bin covering problem, which is known to be NP-hard; see e.g., [2]. However, as we shall see in this paper, the special structure of the factorial (and in particular, the profusion of factors at the "tiny primes" $2, 3$) make it more tractable than the general case.

Remark 1.2. One can equivalently define $t(N)$ as the greatest t for which there exists a t-admissible *subfactorization* of $N!$ of cardinality *at least* N. This is because every such subfactorization can be converted into a t-admissible factorization of cardinality exactly N by first deleting elements from the subfactorization to make the cardinality N, and then multiplying one of the elements of the subfactorization by a natural number to upgrade the subfactorization to a factorization. This "relaxed" formulation of the problem turns out to be more convenient for both theoretical analysis of $t(N)$ and numerical computations.

By combining the obvious lower bound

$$\prod \mathcal{B} \geq t^{|\mathcal{B}|} \tag{1.1}$$

for any t-admissible multiset $\mathcal{B}$ with Stirling's formula (2.11), we obtain the trivial upper bound

$$\frac{t(N)}{N} \leq \frac{(N!)^{1/N}}{N} = \frac{1}{e} + O\left(\frac{\log N}{N}\right) \tag{1.2}$$

for $N \geq 2$; see Figure 1. In [8, p.75] it was reported that an unpublished work of Erdős, Selfridge, and Straus established the asymptotic

$$\frac{t(N)}{N} = \frac{1}{e} + o(1) \tag{1.3}$$

(first conjectured in [6]) and asked if one could show the bound

$$\frac{t(N)}{N} \leq \frac{1}{e} - \frac{c}{\log N} \tag{1.4}$$

for some constant $c > 0$ (problem #391 in `https://www.erdosproblems.com`; see also [9, Section B22, p. 122–123]); it was also noted that similar results were obtained in [1] if one restricted the a_i to be prime powers. However, as later reported in [7], Erdős "believed that Straus had written up our proof [of (1.3)]. Unfortunately Straus suddenly died and no trace was ever found of his notes. Furthermore, we never could reconstruct our proof, so our assertion now can be called only a conjecture". In [9] the lower bound $\frac{t(N)}{N} \geq \frac{1}{4}$ was established for sufficiently large N, by rearranging powers of 2 and 3 in the obvious factorization $1 \times 2 \times \cdots \times N$ of $N!$. A variant lower bound of the asymptotic shape $\frac{t(N)}{N} \geq \frac{3}{16} - o(1)$ obtained by rearranging only powers of 2, and which is superior for medium values of N, can also be found in [9]. The following conjectures in [9] were also made:

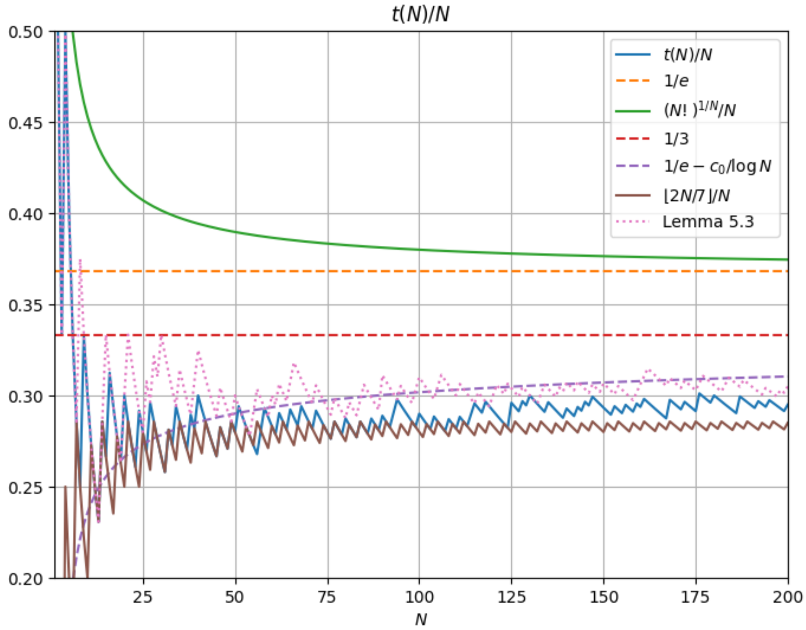

FIGURE 1. The function $t(N)/N$ (blue) for $N \leq 200$, using the data from OEIS A034258, as well as the trivial upper bound $(N!)^{1/N}/N$ (green), the improved upper bound from Lemma 5.3 (pink), which is asymptotic to (1.5) (purple), and the function $\lfloor 2N/7 \rfloor/N$ (brown), which we show to be a lower bound for $N \neq 56$. Theorem 1.3 implies that $t(N)/N$ is asymptotic to (1.5) (purple), which in turn converges to $1/e$ (orange). The threshold $1/3$ (red) is permanently crossed at $N = 43632$.

(1) One has $t(N) \leq N/e$ for $N \neq 1, 2, 4$.
(2) One has $t(N) \geq \lfloor 2N/7 \rfloor$ for $N \neq 56$.
(3) One has $t(N) \geq N/3$ for $N \geq 3 \times 10^5$. (It was also asked if the threshold 3×10^5 could be lowered.)

In this paper we answer all of these questions.

Theorem 1.3 (Main theorem)**.** *Let N be a natural number.*

(i) *If $N \neq 1, 2, 4$, then $t(N) \leq N/e$.*
(ii) *If $N \neq 56$, then $t(N) \geq \lfloor 2N/7 \rfloor$.*
(iii) *If $N \geq 43632$, then $t(N) \geq N/3$. The threshold 43632 is best possible.*
(iv) *For large N, one has*

$$\frac{t(N)}{N} = \frac{1}{e} - \frac{c_0}{\log N} + O\left(\frac{1}{\log^{1+c} N}\right) \tag{1.5}$$

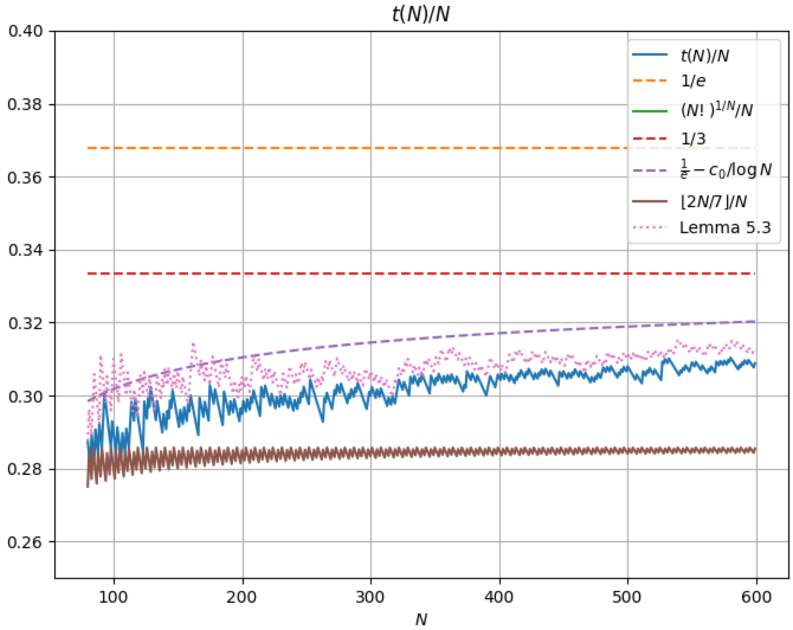

FIGURE 2. A continuation of Figure 1 to the region $80 \leq N \leq 599$.

for some constant $c > 0$*, where* c_0 *is the explicit constant*

$$\begin{aligned} c_0 &:= \frac{1}{e}\int_0^1 f_e(x)\,dx \\ &= 0.3044190\dots \end{aligned} \tag{1.6}$$

and for any $\alpha > 0$*,* $f_\alpha : (0,\infty) \to \mathbb{R}$ *denotes the piecewise smooth function*

$$f_\alpha(x) := \left\lfloor \frac{1}{x} \right\rfloor \log \frac{\lceil 1/\alpha x \rceil}{1/\alpha x}. \tag{1.7}$$

In particular, (1.3) *and* (1.4) *hold.*

For future reference, we observe the simple bounds

$$\begin{aligned} 0 \leq f_\alpha(x) &\leq \frac{1}{x}\log\frac{1/\alpha x + 1}{1/\alpha x} \\ &= \frac{1}{x}\log(1+\alpha x) \\ &\leq \alpha \end{aligned} \tag{1.8}$$

for all $x > 0$; in particular, f_α is a bounded function. It however has an oscillating singularity at $x = 0$; see Figure 3.

In Appendix C we give some details on the numerical computation of the constant c_0.

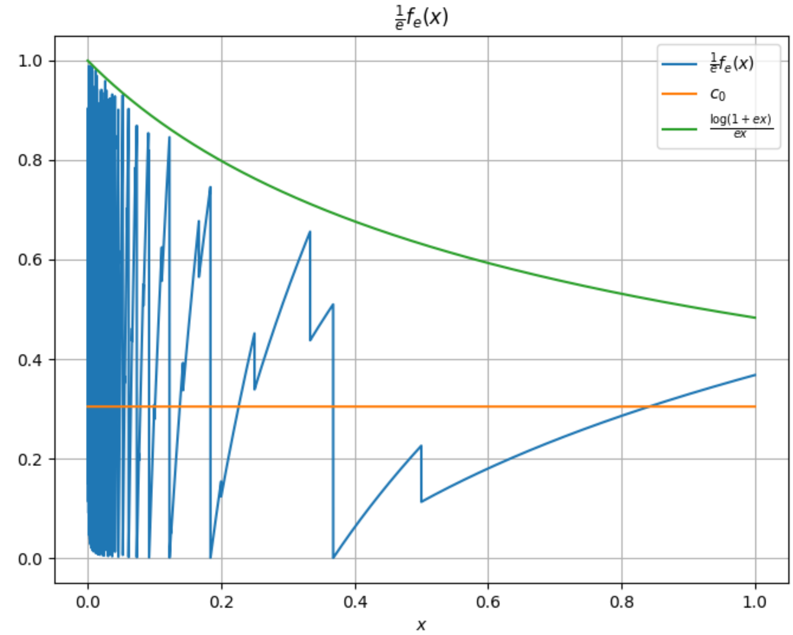

FIGURE 3. The piecewise continuous function $x \mapsto \frac{1}{e} f_e(x)$, together with its mean value $c_0 = 0.3044190\dots$ and the upper bound $\frac{\log(1+ex)}{ex}$. The function exhibits an oscillatory singularity at $x = 0$ similar to $\sin \frac{1}{x}$ (but it is always nonnegative and bounded). Informally, the function f_e quantifies the difficulty that large primes in the factorization of $N!$ have in becoming slightly larger than N/e after multiplying by a natural number.

Remark 1.4. In a previous version [15] of this manuscript, the weaker bounds

$$\frac{1}{e} - \frac{O(1)}{\log N} \leq \frac{t(N)}{N} \leq \frac{1}{e} - \frac{c_0 + o(1)}{\log N}$$

were established, which were enough to recover (1.3), (1.4), and Theorem 1.3(i).

As one might expect, the proof of Theorem 1.3 proceeds by a combination of both theoretical analysis and numerical calculations. Our main tools to obtain upper and lower bounds on $t(N)$ can be summarized as follows:

- In Section 4, we discuss *greedy algorithms* to construct subfactorizations, that provide quickly computable, though suboptimal, lower bounds on $t(N)$ for small and medium values;
- In Section 3, we present a *linear programming* (or *integer programming*) method that provides quite accurate upper and lower bounds on $t(N)$ for small and medium values of N;
- In Section 5, we introduce an *accounting identity* linking the "t-excess" of a subfactorization with its "p-surpluses" at various primes, which provides an reasonable upper bound on $t(N)$ for all N, and is discussed in more detail in Section 5;

- In Section 6, we extend the *rearrangement approach* from [10] to give a computer-assisted proof that Theorem 1.3(iii) holds for sufficiently large N.
- In Section 7, we give *modified approximate factorization* strategy, which provides lower bounds on $t(N)$, that become asymptotically quite efficient.

The final approach is significantly more complicated than the other three, but gives the most efficient lower bounds in the asymptotic limit $N \to \infty$. The key idea is to start with an approximate factorization

$$N! \approx \left(\prod_{j \in I} j\right)^A$$

for some small natural number A (e.g., $A = \lfloor \log^2 N \rfloor$) and a suitable set I of natural numbers greater than or equal to t; there is some freedom to select parameters here, and we will take I to be the natural numbers in $(t, t(1+\sigma)]$ that are coprime to 6, where t is the target lower bound for $t(N)$ we wish to establish, and $\sigma := \frac{3N}{tA}$. With a suitable choice of I, this product contains approximately the right number of copies of p for medium-sized primes p; but it has the "wrong" number of copies of large primes, and is also constructed to avoid the "tiny" primes $p = 2, 3$. One then performs a number of alterations to this approximate factorization to correct for the "surpluses" or "deficits" at various primes $p > 3$, using the supply of available tiny primes $p = 2, 3$ as a sort of "liquidity pool" to efficiently reallocate primes in the factorization. A key point will be that the incommensurability of $\log 2$ and $\log 3$ (i.e., the irrationality of $\log 3/\log 2$) means that the 3-smooth numbers (numbers of the form $2^n 3^m$) are asymptotically dense (in logarithmic scale), allowing for other factors to be exchanged for 3-smooth factors with little loss[1].

1.1. **Author contributions and data.** This project was initially concieved as a single-author manuscript by Terence Tao, but since the release of the initial preprint [15], grew to become a collaborative project organized via the Github repository [16], which also contains the supporting code and data for the project. The contributions of the individual authors, according to the CRediT categories at `https://credit.niso.org/`, are as follows:

authors should be arranged in alphabetical order of surname.

- Boris Alexeev: ...
- ...
- Terence Tao: Conceptualization, Formal Analysis, Methodology, Project Administration, Visualization, Writing – original draft, Writing – review & editing.
- Kevin Ventullo: ...

1.2. **Acknowledgments.** TT is supported by NSF grant DMS-2347850. We thank Thomas Bloom for the web site `https://www.erdosproblems.com`, where the author learned of this problem, as well as Bryna Kra and Ivan Pan for corrections.

[1]The weaker results alluded to in Remark 1.4 only used the prime 2 as a supply of "liquidity", and thus encountered inefficiencies due to the inability to "make change" when approximating another factor by a power of two.

list here all contributors to the project who did not wish to be listed as co-authors.

2. Notation and basic estimates

We use the usual asymptotic notation $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote an inequality of the form $|X| \leq CY$ for some absolute constant C. We also write $X \asymp Y$ for $X \ll Y \ll X$. For effective estimates, we will use the more precise notation $O_{\leq}(Y)$ to denote any quantity whose magnitude is bounded by exactly at most Y. We also use $O_{\leq}(Y)^+$ to denote a quantity of size $O_{\leq}(Y)$ that is in addition non-negative, that is to say it lies in the interval $[0, Y]$.

If S is a statement, we use 1_S to denote its indicator, thus $1_S = 1$ when S is true and $1_S = 0$ when S is false. If x is a real number, we use $\lfloor x \rfloor$ to denote the greatest integer less than or equal to x, and $\lceil x \rceil$ to be the least integer greater than or equal to x.

Throughout this paper, the symbol p (or p', p_1, p_2, etc.) is always understood to be restricted to be prime. The primes $2, 3$ will play a special role in this paper and will be referred to as *tiny primes*. Call a natural number 3-*smooth* if it is the product of tiny primes, i.e., it is of the form $2^n 3^m$ for some natural numbers n, m, and 3-*rough* if it is not divisible by any tiny prime, that is to say it is coprime to 6. Given a positive real number x, we use $\lceil x \rceil^{\langle 2,3 \rangle}$ to denote the smallest 3-smooth number greater than or equal to x. For instance, $\lceil 5 \rceil^{\langle 2,3 \rangle} = 6$ and $\lceil 10 \rceil^{\langle 2,3 \rangle} = 12$.

It will be convenient to introduce a variant of this quantity that is close to a power of 12. If $1 \leq L \leq x$ is an additional real parameter, we define

$$\lceil x \rceil_L^{\langle 2,3 \rangle} := 12^a \lceil x/12^a \rceil^{\langle 2,3 \rangle} \tag{2.1}$$

for any real $x \geq L \geq 1$, where $a := \lfloor \frac{x/L}{\log 12} \rfloor$ is the largest integer such that $12^a \leq x/L$.

For any $L \geq 1$, let κ_L be the least quantity such that

$$x \leq \lceil x \rceil^{\langle 2,3 \rangle} \leq \exp(\kappa_L) x \tag{2.2}$$

holds for all $x \geq L$; see Figure 4. In Appendix A we establish the following facts:

Lemma 2.1 (Approximation by 3-smooth numbers)**.**

(i) *We have* $\kappa_{4.5} = \log \frac{4}{3} = 0.28768\dots$ *and* $\kappa_{40.5} = \log \frac{32}{27} = 0.16989\dots$.
(ii) *For large* L, *one has* $\kappa_L \ll \log^{-c} L$ *for some absolute constant* $c > 0$.
(iii) *If* $1 \leq L \leq x$ *are real numbers, then*

$$x \leq \lceil x \rceil_L^{\langle 2,3 \rangle} \leq \exp(\kappa_L) x \tag{2.3}$$

and for any $0 \leq \gamma < 1$ *we have*

$$\frac{\nu_2(\lceil x \rceil_L^{\langle 2,3 \rangle}) - 2\gamma \nu_3(\lceil x \rceil_L^{\langle 2,3 \rangle})}{1-\gamma} \leq \frac{2}{\log 12}\left(\log x + \kappa_{L,\gamma}^{(2)}\right) \tag{2.4}$$

and

$$\frac{2\nu_3(\lceil x \rceil_L^{\langle 2,3 \rangle}) - \gamma \nu_2(\lceil x \rceil_L^{\langle 2,3 \rangle})}{1-\gamma} \leq \frac{2}{\log 12}\left(\log x + \kappa_{L,\gamma}^{(3)}\right) \tag{2.5}$$

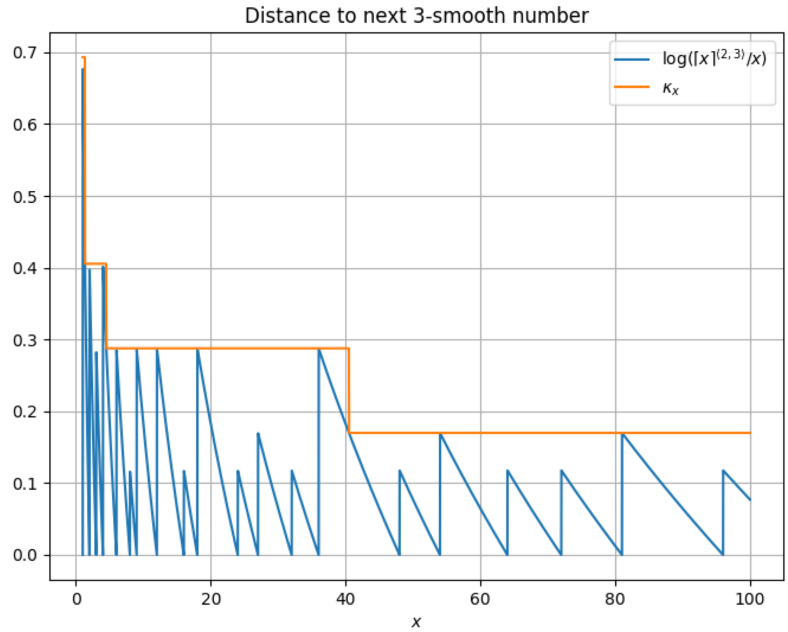

FIGURE 4. The function $\log \lceil x \rceil^{\langle 2,3 \rangle}/x$, compared against κ_x.

where

$$\kappa^{(2)}_{L,\gamma} := \left(\frac{\log 12}{2(1-\gamma)\log 2} - 1 \right) \log(12L) + \frac{\kappa_L \log 12}{2(1-\gamma)\log 2} \tag{2.6}$$

$$\kappa^{(3)}_{L,\gamma} := \left(\frac{\log 12}{(1-\gamma)\log 3} - 1 \right) \log(12L) + \frac{\kappa_L \log 12}{(1-\gamma)\log 3}. \tag{2.7}$$

We remark that when x is a power of 12, the left-hand sides of (2.4), (2.5) are both equal to $\frac{2}{\log 12} \log x$; thus the estimates (2.4), (2.5) are quite efficient asymptotically.

We use (a, b) to denote the greatest common divisor of a and b, $a|b$ to denote the assertion that a divides b, and $\pi(x) = \sum_{p \leq x} 1$ to denote the usual prime counting function.

We use $v_p(a/b) = v_p(a) - v_p(b)$ to denote the p-adic valuation of a positive natural number a/b, that is to say the number of times p divides the numerator a, minus the number of times p divides the denominator b. For instance, $v_2(32/27) = 5$ and $v_3(32/27) = -3$. If one applies a logarithm to the fundamental theorem of arithmetic, one obtains the identity

$$\sum_p v_p(r) \log p = \log r \tag{2.8}$$

for any positive rational r.

For a natural number n, we can write

$$\nu_p(n) = \sum_{j=1}^{\infty} 1_{p^j | n}. \tag{2.9}$$

Upon taking partial sums, we recover Legendre's formula

$$\nu_p(N!) = \sum_{j=1}^{\infty} \left\lfloor \frac{N}{p^j} \right\rfloor = \frac{N - s_p(N)}{p-1} \tag{2.10}$$

where $s_p(N)$ is the sum of the digits of N in the base p expansion.

Given a putative factorization $\mathcal{B}$ of $N!$, we refer to the quantity $\nu_p\left(\frac{N!}{\prod \mathcal{B}}\right)$ as the *p-surplus* of $\mathcal{B}$ with respect to the target $N!$; if it is negative, we refer to $-\nu_p\left(\frac{N!}{\prod \mathcal{B}}\right) = \nu_p\left(\frac{\prod \mathcal{B}}{N!}\right)$ as the *p-deficit*, with the multiset being *p-balanced* if the p-surplus (or p-deficit) is zero. Thus, a factorization of $N!$ is achieved if and only if one is balanced at every prime p, whereas a subfactorization is achieved if one is either in balance or surplus at every prime p.

To bound the factorial, we have the explicit Stirling approximation [13]

$$\log N! = N \log N - N + \log \sqrt{2\pi N} + O^{+}_{\le}\left(\frac{1}{12N}\right), \tag{2.11}$$

valid for all natural numbers N.

We recall the effective prime number theorem from [5, Corollary 5.2], which asserts that

$$\pi(x) \ge \frac{x}{\log x} + \frac{x}{\log^2 x} \tag{2.12}$$

for $x \ge 599$ and

$$\pi(x) \le \frac{x}{\log x} + \frac{1.2762x}{\log^2 x} \tag{2.13}$$

for $x > 1$.

We will also need to control sums of somewhat oscillatory functions over primes, for which the bounds in (2.12), (2.13) are of insufficient strength. Let $y < x$ be real numbers. Given a function $b\colon (y,x] \to \mathbb{R}$, its *total variation* $\|b\|_{\mathrm{TV}(y,x]}$ is defined as the supremum of the quantities $\sum_{j=0}^{J-1} |b(x_{j+1}) - b(x_j)|$ for $y < x_0 \le \cdots \le x_J \le x$, and the *augmented total variation* $\|b\|_{\mathrm{TV}^*(y,x]}$ is defined as

$$\|b\|_{\mathrm{TV}^*(y,x]} := |b(y^+)| + |b(x)| + \|b\|_{\mathrm{TV}(y,x]},$$

$b(y^+) := \lim_{t\to y^+} b(t)$ denotes the right limit of b at y (if it exists). Equivalently, $\|b\|_{\mathrm{TV}^*(y,x]}$ is the total variation of b if extended by zero outside of $(y,x]$. The indicator function $1_{(y,x]}$ clearly has an augumented total variation of 2.

We will use this augmented total variation to control sums over primes. More precisely, in Appendix B we will show

Lemma 2.2 (Effective bounds for oscillatory sums over primes)**.** *Let* $1423 \leq y \leq x$*, and let* $b : (y, x] \to \mathbb{R}$ *be of bounded total variation. Then we have the bound*

$$\sum_{y<p\leq x} b(p) \log p = \int_y^x \left(1 - \frac{2}{\sqrt{t}}\right) b(t)\, dt + O_{\leq}(\|b\|_{\mathrm{TV}^*(y,x]} E(x)) \tag{2.14}$$

where the error function $E(x)$ *is defined as*

$$E(x) := 0.95\sqrt{x} + 3.83 \times 10^{-9} x. \tag{2.15}$$

In particular one has

$$\pi(x) - \pi(y) = \int_y^x (1 - \frac{2}{\sqrt{t}}) \frac{dt}{\log t} + O_{\leq}(2\frac{E(x)}{\log y}). \tag{2.16}$$

If b is non-negative, one also has the upper bound

$$\sum_{y<p\leq x} b(p) \leq \frac{1}{\log y} \int_y^x b(t)\, dt + \|b\|_{\mathrm{TV}^*(y,x]} \frac{E(x)}{\log y} \tag{2.17}$$

and the lower bound

$$\sum_{y<p\leq x} b(p) \leq \frac{1 - \frac{2}{\sqrt{y}}}{\log x} \int_y^x b(t)\, dt - \|b\|_{\mathrm{TV}^*(y,x]} \frac{E(x)}{\log x}. \tag{2.18}$$

Thus for instance

$$\pi(x) - \pi(y) \leq \frac{x - y}{\log y} + 2\frac{E(x)}{\log y} \tag{2.19}$$

and

$$\pi(x) - \pi(y) \geq \left(1 - \frac{2}{\sqrt{y}}\right) \frac{x - y}{\log x} - 2\frac{E(x)}{\log x}. \tag{2.20}$$

One can also replace all occurrences of $E(x)$ *here by the classical error term* $O(x \exp(-c\sqrt{\log x}))$ *for some absolute constant* $c > 0$*.*

We remark that the accuracy in (2.14), (2.16) in particular is on par with what would be provided by the Riemann hypothesis, as long as x is not too large (e.g., $x \leq 10^{16}$). The other estimates are not quite as precise, but still adequate for our applications. The error term $E(x)$ can be improved somewhat for large x (see (B.3)), but this simplified version will suffice for our analysis (in particular, the contribution of the second term in (2.15) will be negligible for our applications).

3. Linear programming

A surprisingly sharp upper bound on $t(N)$ comes from linear programming.

Lemma 3.1 (Linear programming bound)**.** *Let N be an natural number and $1 \leq t \leq N/2$. Suppose for each prime $p \leq N$, one has a non-negative real number w_p which is weakly non-decreasing in p (thus $w_p \leq w_{p'}$ when $p \leq p'$), and such that*

$$\sum_p w_p \nu_p(j) \geq 1 \tag{3.1}$$

for all $t \leq j \leq N$, and such that

$$\sum_p w_p \nu_p(N!) < N. \tag{3.2}$$

Then $t(N) < t$.

Proof. We first observe that the bound (3.1) in fact holds for all $j \geq t$, not just for $t \leq j \leq N$. Indeed, if this were not the case, consider the first $j \geq t$ where (3.1) fails. Take a prime p dividing j and replace it by a prime in the interval $[p/2, p)$ which exists by Bertrand's postulate (or remove p entirely, if $p = 2$); this creates a new j' in $[j/2, j)$ which is still at least t. By the weakly decerasing hypothesis on w_p, we have

$$\sum_p w_p \nu_p(j) \geq \sum_p w_p \nu_p(j')$$

and hence by the minimality of j we have

$$\sum_p w_p \nu_p(j) > 1,$$

a contradiction.

Now suppose for contradiction that $t(N) \geq t$, thus we have a factorization $N! = \prod_{j \geq t} j^{m_j}$ for some natural numbers m_j summing to N. Taking p-valuations, we conclude that

$$\sum_{j \geq t} m_j \nu_p(j) \leq \nu_p(N!)$$

for all $p \leq N$. Multiplying by w_p and summing, we conclude from (3.1) that

$$N = \sum_{j \geq t} m_j \leq \sum_p w_p \nu_p(N!),$$

contradicting (3.2). □

This bound is sharp for all $N \leq 600$, with the exception of $N = 155$, where it gives the upper bound $t(155) \leq 46$. A more precise integer program (discussed below) gives $t(155) = 45$.

A variant of the linear programming method also gives good lower bound constructions. Specifically, one can use linear programming to find non-negative real numbers m_j for $t \leq j \leq N$ that maximize the quantity $\sum_{t \leq j \leq N} m_j$ subject to the constraints

$$\sum_{t \leq j \leq N} m_j \nu_p(j) \leq \nu_p(N!).$$

The expression $\prod_{t \leq j \leq N} j^{\lfloor m_j \rfloor}$ will then be a subfactorization of $N!$ into $\sum_{t \leq j \leq N} \lfloor m_j \rfloor$ factors j, each of which is at least t. If $\sum_{t \leq j \leq N} \lfloor m_j \rfloor \geq N$, this demonstrates that $t(N) \geq t$. Numerically,

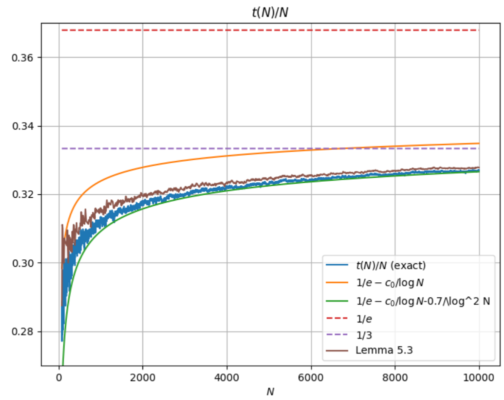

FIGURE 5. $t(N)/N$ for $80 \leq N \leq 10^4$, obtained via linear programming in most cases (and integer programming in some exceptional cases). The upper bound from Lemma 5.3 is surprisingly sharp; on the other hand, the asymptotic $1/e - c_0/\log N$ is a relatively poor approximation, suggesting the presence of a lower order term that empirically appears to be about $1/e - c_0/\log N - 0.7/\log^2 N$.

this procedure attains the exact value of $t(N)$ for all $N \leq 600$; for instance for $N = 155$, it shows that $t(155) \geq 45$.

discuss integer programming, need to restrict j to a finite set of "useful" integers

These methods also give quite precise upper and lower bounds for larger values of N, but with quite slow runtime. For instance, with $N = 3 \times 10^5$ and $t = N/3 = 10^5$, the upper bound method can be used to show that any t-admissible factorization has cardinality at most $N + 455$, while the lower bound method produces a t-admissible factorization of exactly this cardinality.

more discussion here

By using the greedy method, Theorem 1.3(ii) can be verified for $N \leq 3 \times 10^5$, and Theorem 1.3(iii) can be verified for $8 \times 10^4 \leq N \leq ???$. The linear programming method can also estabalish Theorem 1.3(iii) in the range $43632 \leq N \leq 8 \times 10^4$. Thus, to resolve these claims, it remains to only establish Theorem 1.3(iii) in the regime $N > ???$.

4. GREEDY ALGORITHMS

The following simple greedy algorithm gives reasonably good performance to obtain large t-admissible subfactorizations $\mathcal{B}$ of $N!$ for a given choice of t and N:

(0) Initialize $\mathcal{B}$ to be the empty multiset.
(1) If $\mathcal{B}$ is not a factorization, locate the largest prime p which is currently in surplus: $\nu_p(N!/\prod \mathcal{B}) > 0$.

(2) If $N!/\prod \mathcal{B}$ contains a multiple of p that is greater than or equal to t, locate the smallest such multiple, add it to $\mathcal{B}$, and return to Step 1. Otherwise, `HALT` the algorithm.

This procedure clearly halts in finite time to produce a t-admissible subfactoriation of $N!$. For instance, applying this procedure with $N = 9$, $t = 3$ produces the 3-admissible subfactorization

$$\{7\times 1, 5\times 1, 3\times 1, 3\times 1, 3\times 1, 3\times 1, 2\times 2, 2\times 2, 2\times 2\}$$

which recovers the bound $t(9) \geq 3$ from Example 1.1 (though with a slightly different subfactorization, in which the 8 is replaced by 4).

This procedure is efficient for small N, for instance attaining the exact value of $t(N)$ for all $N \leq 79$, though it begins to degrade for larger N; see Figure 6. The performance is also respectable (though not optimal) for medium N; for instance, when $N = 3\times 10^5$ and $t = N/3$, it locates a t-admissible subfactorization of $N!$ of cardinality $N + 372$, which is close to the linear programming limit of $N + 455$.

discuss modifications to the algorithm to make it perform both faster and more accurately

5. The accounting identity

Given a t-admissible multiset $\mathcal{B}$ (which we view as an approximate factorization of $N!$), we can apply the fundamental theorem of arithmetic (2.8) to the rational number $N!/\prod \mathcal{B}$ and rearrange to obtain the *accounting identity*

$$\mathcal{E}_t(\mathcal{B}) + \sum_p \nu_p\left(\frac{N!}{\prod \mathcal{B}}\right)\log p = \log N! - |\mathcal{B}|\log t \tag{5.1}$$

where we define the *t-excess* $\mathcal{E}_t(\mathcal{B})$ of the multiset $\mathcal{B}$ by the formula

$$\mathcal{E}_t(\mathcal{B}) := \sum_{a\in\mathcal{B}} \log\frac{a}{t}. \tag{5.2}$$

Example 5.1. Suppose one wishes to factorize $5! = 2^3 \times 3 \times 5$. The attempted 3-admissible factorization $\mathcal{B} := \{3, 4, 5, 5\}$ has a 2-surplus of $\nu_2(5!/\prod \mathcal{B}) = 1$, is in balance at 3, and has a 5-deficit of $\nu_2(\prod \mathcal{B}/5!) = 1$, so it is not a factorization or subfactorization of $5!$. The 3-excess of this multiset is

$$\mathcal{E}_3(\mathcal{B}) = \log\frac{3}{3} + \log\frac{4}{3} + \log\frac{5}{3} + \log\frac{5}{3} = 1.3093\dots$$

and the accounting identity (5.1) become

$$1.3093\cdots + \log 2 - \log 5 = 0.3930\cdots = \log 5! - 4\log 3.$$

If one replaces one of the copies of 5 in $\mathcal{B}$ with a 2, this erases both the 2-surplus and the 5-deficit, and creates a factorization $\mathcal{B}' = \{2, 3, 4, 5\}$ of $5!$; the 3-excess now drops to

$$\mathcal{E}_3(\mathcal{B}) = \log\frac{2}{3} + \log\frac{3}{3} + \log\frac{4}{3} + \log\frac{5}{3} = 0.3930\dots,$$

bringing the accounting identity back into balance.

In view of Remark 1.2, one can now equivalently describe $t(N)$ as follows:

Lemma 5.2 (Equivalent description of $t(N)$)**.** *$t(N)$ is the largest quantity t for which there exists a t-admissible subfactorization of $N!$ with*

$$\mathcal{E}_t(\mathcal{B}) + \sum_p \nu_p\left(\frac{N!}{\prod \mathcal{B}}\right)\log p \le \log N! - N\log t.$$

One can view $\log N! - N\log t$ as an available "budget" that one can "spend" on some combination of t-excess and p-surpluses. For t of the form $t = N/e^{1+\delta}$ for some $\delta > 0$, the budget can be computed using the Stirling approximation (2.11) to be $\delta N + O(\log N)$. The non-negativity of the t-excess and p-surpluses recovers the trivial upper bound (1.2); but one can improve upon this bound by observing that large prime factors of $N!$ inevitably generate a noticeable t-excess, as follows.

Lemma 5.3 (Upper bound criterion)**.** *Suppose that $1 \le t \le N$ are such that*

$$\sum_{\frac{t}{\lfloor\sqrt{t}\rfloor} < p \le N} f_{N/t}(p/N) > \log N! - N\log t, \tag{5.3}$$

where $f_{N/t}$ was defined in (1.7)*. Then $t(N) < t$.*

Proof. Suppose for contradiction that $t(N) \ge t$, then we can find a t-admissible factorization $\mathcal{B}$ of $N!$. The accounting identity then gives

$$\sum_{a\in\mathcal{B}} \log\frac{a}{t} = \mathcal{E}_t(\mathcal{B}) = \log N! - N\log t. \tag{5.4}$$

We write $f_{N/t}(p/N) = \lfloor \frac{N}{p}\rfloor g_t(p)$, where $g_t(p) := \log(\frac{p}{t}\lceil\frac{t}{p}\rceil)$. We claim that

$$\log\frac{a}{t} \ge g_t(p_{a,1}) + \cdots + g_t(p_{a,k_a}) \tag{5.5}$$

for all $a \in \mathcal{B}$, where $p_{a,1}, \dots, p_{a,k_a}$ are the primes greater than $\frac{t}{\lfloor\sqrt{t}\rfloor}$ that divide a (counting multiplicity). For $k_a = 0$ this is clear since $a \ge t$. For $k_a = 1$, we can write $a = d_a p_{a,1}$ where $p_{a,1} > \frac{t}{\sqrt{t}+1}$ and $d_a \ge \lceil\frac{t}{p_{a,1}}\rceil$, so that

$$\log\frac{a}{t} = \log\left(\frac{p_{a,1}}{t}d_a\right) \ge g_t(p_{a,1}),$$

again giving (5.5). For $k_a \ge 2$, we have $a \ge p_{a,1}\dots p_{a,k}$, hence

$$\begin{aligned}
\log\frac{a}{t} - \sum_{j=1}^{k_a} g_t(p_{a,j}) &\ge \sum_{j=1}^{k_a}(\log p_{a,j} - g_t(p_{a,j})) - \log t \\
&= \sum_{j=1}^{k_a}\left(\log t - \log\left\lceil\frac{t}{p_{a,j}}\right\rceil\right) - \log t \\
&\ge \sum_{j=1}^{k_a}\left(\log t - \log\sqrt{t}\right) - \log t \\
&\ge 0
\end{aligned}$$

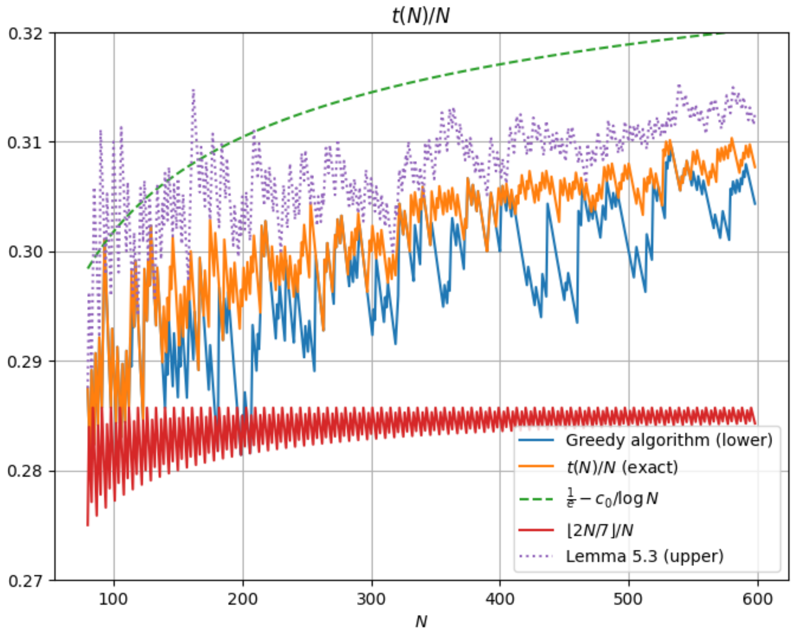

FIGURE 6. An enlarged version of Figure 2, displaying the lower bound from the greedy algorithm and the upper bound from Lemma 5.3. The linear programming upper and lower bounds are exact in this region, except for $N = 155$ in which the upper bound is off by one.

which again gives (5.5). Summing (5.5) over all $a \in \mathcal{B}$ and inserting into (5.4), we conclude that

$$\sum_{p > \frac{t}{\lfloor \sqrt{t} \rfloor}} v_p(N!) g_t(p) \leq \log N! - N \log t.$$

By (2.10), we can bound $v_p(N!) g_t(p)$ by $\lfloor N/p \rfloor g_t(p) = f_{N/t}(p/N)$. This contradicts (5.3), giving the claim. $\square$

In practice, Lemma 5.3 gives reasonable upper bounds on N, especially when N is large, although for medium N the linear programming method is superior: see Figure 1, Figure 2, Figure 6.

We can now prove the upper bound portion of Theorem 1.3(iv):

Proposition 5.4. *For large N, one has*

$$\frac{t(N)}{N} \leq \frac{1}{e} - \frac{c_0}{\log N} + O\left(\frac{1}{\log^2 N}\right).$$

Proof. We apply Lemma 5.3 with

$$t := \frac{1}{e} - \frac{c_0}{\log N} + \frac{C_0}{\log^2 N}$$

with C_0 a large absolute constant to be chosen later. From Taylor expansion and the Stirling approximation one sees that

$$\log N! - N \log t \geq ec_0 \frac{N}{\log N} + (C_0 - O(1)) \frac{N}{\log^2 N}$$

so it will suffice to establish the upper bound

$$\sum_{\frac{t}{\lfloor \sqrt{t} \rfloor} < p \leq N} f_{N/t}(p/N) \leq ec_0 \frac{N}{\log N} + O\left(\frac{N}{\log^2 N}\right).$$

For N large enough, we have $\frac{t}{\lfloor \sqrt{t} \rfloor} \leq \frac{N}{\log N}$, so it suffices to show that

$$\sum_{\frac{N}{\log N} \leq p \leq N} f_{N/t}(p/N) \leq ec_0 \frac{N}{\log N} + O\left(\frac{N}{\log^2 N}\right).$$

On the interval $[1/\log N, 1]$, the piecewise smooth function $f_{N/t}$ is bounded by $O(1)$ thanks to (1.8), and has a total variation of $O(\log N)$; the same is then true for the rescaled function $x \mapsto f_{N/t}(x/N)$ on $[N/\log N, 1]$. By Lemma 2.2 (with classical error term), the left-hand side is then

$$\int_{N/\log N}^{N} \left(1 - \frac{2}{\sqrt{x}}\right) f_{N/t}(x/N) \frac{dx}{\log x} + O\left(N \exp(-c\sqrt{\log N})\right)$$

for some $c > 0$. Discarding the $\frac{2}{\sqrt{x}}$ term, performing a change of variable, and using (1.6), we reduce to showing that

$$\int_{1/\log N}^{1} f_{N/t}(x) \frac{\log N}{\log(Nx)}\, dx \leq \int_0^1 f_e(x)\, dx + O\left(\frac{1}{\log N}\right).$$

We have the Taylor approximation

$$\frac{\log N}{\log(Nx)} = 1 + O\left(\frac{\log(1/x)}{\log N}\right).$$

Applying (1.8) and the integrability of $\log(1/x)$, we see that the contribution of the error term is acceptable. Applying a further rescaling by $N/et = 1 + O(1/\log N)$, we reduce to showing that

$$\int_{N/et \log N}^{N/et} f_{N/t}(Nx/et)\, dx = \int_0^1 f_e(x)\, dx + O\left(\frac{1}{\log N}\right).$$

But observe that $f_{N/t}(Nx/et) = f_e(x)$ unless $\frac{1}{x}$ is within $O(1/\log N)$ of an integer, which one can calculate to occur on a set of measure $O(1/\log N)$ for $x \in [0, N/et]$. By (1.8), both integrands are bounded by $O(1)$ for all $x \in [0, N/et]$, and the claim follows from the triangle inequality. □

We can now establish Theorem 1.3(i):

Proposition 5.5. *One has $t(N)/N < 1/e$ for $N \neq 1, 2, 4$.*

Proof. From existing data on $t(N)$ (or the linear programming method) one can verify this claim for $N < 80$ (see Figure 1), so we assume that $N \geq 80$.

Applying Lemma 5.3, (2.11), it suffices to show that

$$\sum_{p \geq \frac{N/e}{\lfloor \sqrt{N/e} \rfloor}} f_e(p/N) > \frac{1}{2}\log(2\pi N) + \frac{1}{12N}. \tag{5.6}$$

This may be easily verified numerically in the range $80 \leq N \leq 5000$ (see Figure 7). We will discard the $\lfloor \sqrt{N/e} \rfloor$ denominator, and reduce to showing

$$\sum_{N/e < p \leq N} f_e(p/N) > \frac{1}{2}\log(2\pi N) + \frac{1}{12N} \tag{5.7}$$

for $N > 5000$. On $[1/e, 1]$, one can compute

$$\|f_e\|_{\mathrm{TV}^*(1/e,1]} = 4 - 2\log 2$$

so by Lemma 2.2 (noting that $5000 > 1423e$) we have

$$\sum_{N/e < p \leq N} f_e(p/N) \log p \geq N \int_{1/e}^{1} \left(1 - \frac{2}{\sqrt{Nx}}\right) f_e(x)\, dx - (4 - 2\log 2) E(N).$$

By upper bounding $\log p$ by $\log N$ and lower bounding $\left(1 - \frac{2}{\sqrt{Nx}}\right)$ by $1 - \frac{2}{\sqrt{N/e}}$, it suffices to show that

$$\left(1 - \frac{2}{\sqrt{N/e}}\right) \int_{1/e}^{1} f_e(x)\, dx \geq (4 - 2\log 2)\frac{E(N)}{N} + \frac{\log N \log(2\pi N)}{2N} + \frac{\log N}{12N^2},$$

which is easily verified for $N \geq 5000$ (one has $\int_{1/e}^{1} f_e(x)\, dx = \frac{2}{e} - \frac{\log 2}{2} = 0.3891\ldots$ and $4 - 2\log 2 = 2.613\ldots$, while $E(N)/N \leq 0.015$, and the other two terms on the right-hand side are negligible). □

6. Rearranging the standard factorization

In this section we describe an approach to establishing lower bounds on $t(N)$ by starting with the standard factorization $\{1, \ldots, N\}$, dividing out some small prime factors from some of the terms, and then redistributing them to other terms. This approach was introduced in [9] to give lower bounds of the shape $\frac{t(N)}{N} \geq \frac{3}{16} + o(1)$ (by redistributing powers of two only) and $\frac{t(N)}{N} \geq \frac{1}{4} + o(1)$. With computer assistance, we are also able to show that $\frac{t(N)}{N} \geq \frac{1}{3} + o(1)$ for sufficiently large N, in a simpler fashion than the method used to prove Theorem 1.3(iv) in the next section.

details needed here

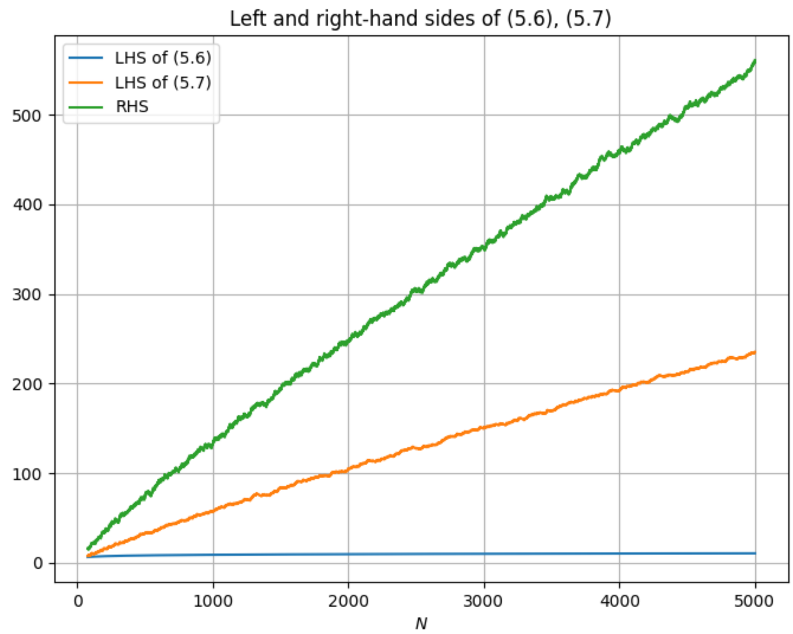

FIGURE 7. A plot of the left and right-hand sides of (5.6), (5.7) for $80 \leq N < 5000$.

7. MODIFIED APPROXIMATE FACTORIZATIONS

In this section we present and then analyze an algorithm that starts with an *approximate* factorization $\mathcal{B}^{(0)}$ of $N!$, which is t-admissible but omits all tiny primes, and is approximately in balance in small and medium primes, and attempts to "repair" this factorization to establish a lower bound of the form $t(N) \geq t$.

To describe the criterion for the algorithm to succeed, it will be convenient to introduce the following notation. For $a_+, a_- \in [0, +\infty]$, we define the asymmetric norm $|x|_{a_+,a_-}$ of a real number x by the formula

$$|x|_{a_+,a_-} := \begin{cases} a_+|x| & x \geq 0 \\ a_-|x| & x \leq 0, \end{cases}$$

with the usual convention $+\infty \times 0 = 0$. If a_+, a_- are finite, this function is Lipschitz with constant $\max(a_+, a_-)$. One can think of a_+ as the "cost" of making x positive, and a_- as the "cost" of making x negative.

The analysis of the algorithm is now captured by the following proposition.

Proposition 7.1 (Repairing an approximate factorization)**.** *Let N, K be natural numbers, and let $1 \leq t \leq N$ be an additional parameter obeying the conditions*

$$\frac{t}{K} \geq \sqrt{N}; \quad \frac{t}{K^2} \geq K \geq 5. \tag{7.1}$$

We also assume that there are additional parameters $\kappa_* > 0$ *and* $0 \leq \gamma_2, \gamma_3 < 1$*, such that there exist* 3*-smooth numbers*

$$t \leq 2^{n_2}3^{m_2}, 2^{n_3}3^{m_3} \leq e^{\kappa_*}t \tag{7.2}$$

such that

$$2m_2 \leq \gamma_2 n_2; \quad n_3 \leq 2\gamma_3 m_3. \tag{7.3}$$

We define the "norm"

$$\|(n,m)\|_\gamma := \max\left(\frac{n-2\gamma_2 m}{1-\gamma_2}, \frac{2m-\gamma_3 n}{1-\gamma_3}\right).$$

Let $\mathcal{B}^{(0)}$ *be a t-admissible multiset of natural numbers, with all elements of* $\mathcal{B}^{(0)}$ *at most* $(t/K)^2$*, and suppose that one has the inequalities*

$$\sum_{i=1}^{8} \delta_i \leq \delta \tag{7.4}$$

and

$$\sum_{i=1}^{7} \alpha_i \leq 1 \tag{7.5}$$

where

$$\delta_1 := \frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(1)}) \tag{7.6}$$

$$\delta_2 := \frac{1}{N}\sum_{t/K<p\leq N} f_{N/t}(p/N) \tag{7.7}$$

$$\delta_3 := \frac{\kappa_{4.5}}{N}\sum_{3<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)\right| \tag{7.8}$$

$$\delta_4 := \kappa_{4.5}\sum_{K<p_1\leq t/K} A_{p_1} \tag{7.9}$$

$$\delta_5 := \kappa_{4.5}\sum_{3<p_1\leq K} |A_{p_1} - B_{p_1}|_{\frac{\log p_1}{\log(t/K^2)},1} \tag{7.10}$$

$$\delta_6 := \frac{\kappa_{4.5}}{N} \tag{7.11}$$

$$\delta_7 := \frac{\kappa_*}{\log t}\left(\frac{\log 12}{2} - B_2\log 2 - B_3\log 3\right) \tag{7.12}$$

$$\delta_8 := \frac{2(\log t + \kappa_*)}{N} \tag{7.13}$$

$$\delta := \frac{1}{N}\log N! - \log t \tag{7.14}$$

$$\alpha_1 := \frac{1}{N} \left\| \left(\nu_2 \left(\prod \mathcal{B}^{(0)} \right), \nu_3 \left(\prod \mathcal{B}^{(0)} \right) \right) \right\|_\gamma \tag{7.15}$$

$$\alpha_2 := \|(B_2, B_3)\|_\gamma \tag{7.16}$$

$$\alpha_3 := \frac{2}{N \log 12} \left(\log \frac{t}{K} + \kappa_{**} \right) \sum_{3 < p_1 \leq t/K} \left| \nu_{p_1} \left(\frac{N!}{\prod \mathcal{B}^{(0)}} \right) \right| \tag{7.17}$$

$$\alpha_4 := \frac{2}{\log 12} \sum_{K < p_1 \leq t/K} \left(\log \frac{t}{p_1} + \kappa_{**} \right) A_{p_1} \tag{7.18}$$

$$\alpha_5 := \frac{2}{\log 12} \sum_{3 < p_1 \leq K} \left| A_{p_1} - B_{p_1} \right|_{\frac{\log p_1}{\log(t/K^2)}(\log K^2 + \kappa_{**}), \log p_1 + \kappa_{**}} \tag{7.19}$$

$$\alpha_6 := \frac{2}{N \log 12} (\log t + \kappa_{**}) \tag{7.20}$$

$$\alpha_7 := \max \left(\frac{\log(2N)}{(1-\gamma_2) N \log 2}, \frac{2 \log(3N)}{(1-\gamma_3) N \log 3} \right) \tag{7.21}$$

$$\kappa_{**} := \max(\kappa^{(2)}_{4.5,\gamma_2}, \kappa^{(3)}_{4.5,\gamma_3}) \tag{7.22}$$

$$A_{p_1} := \frac{1}{N} \sum_m \nu_{p_1}(m) |\{ a \in \mathcal{B}^{(0)} : a = mp \text{ for a prime } p > t/K \}| \tag{7.23}$$

$$B_{p_1} := \frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \left\lfloor \frac{N}{p} \right\rfloor, \tag{7.24}$$

with the convention that the upper bound $p < \frac{t}{m-1}$ in (7.24) *is vacuous when $m = 1$. Then $t(N) \geq t$.*

Note that the quantities δ_2, δ already made an appearance in Lemma 5.3. As we shall see, when N is large we will be able to construct multisets $\mathcal{B}^{(0)}$ that make all the other quantities $\delta_i, i \neq 2$ and α_i quite small, which is how we shall establish such results as Theorem 1.3(iii) and the lower bound in Theorem 1.3(iv).

In practice, the parameter K will be quite small compared to N, and the quantities $\gamma_2, \gamma_3, \kappa_*$ will also be somewhat smaller than 1.

The rest of this section will be devoted to the proof of this proposition. It will be convenient to divide the primes into four classes:

- *Tiny primes* $p = 2, 3$.
- *Small primes* $3 < p \leq K$.
- *Medium primes* $K < p \leq t/K$.
- *Large primes* $p > t/K$.

Initially, the multiset $\mathcal{B}^{(0)}$ may have the "wrong" number of factors at large primes. We fix this by applying the following modifications to $\mathcal{B}^{(0)}$:

(a) Remove all elements of $\mathcal{B}^{(0)}$ that are divisible by a large prime $p > t/K$ from the multiset.
(b) For each large prime $p > t/K$, add $\nu_p(N!)$ copies of $p\lceil t/p\rceil$ to the multiset.

We let $\mathcal{B}^{(1)}$ be the multiset formed after completing both Step (a) and Step (b). We make two simple observations:

- Since the elements of $\mathcal{B}^{(0)}$ are at most $(t/K)^2$, all the elements removed in Step (a) are of the form mp where $m \leq t/K$.
- For each large prime p considered in Step (b), one has $\nu_p(N!) = \lfloor N/p\rfloor$ by (2.10) and (7.1), while $\lceil t/p\rceil \leq K \leq t/K$ (again by (7.1)).

From this, we see that $\mathcal{B}^{(1)}$ is automatically t-admissible, and in balance at any large prime $p > t/K$:

$$\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) = 0.$$

For medium primes $K < p_1 \leq t/K$, one can have some increase in the p_1-surplus coming from Step (a), which is described by (7.23):

$$\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) = \nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right) + NA_{p_1}.$$

For small or tiny primes $p \leq K$, one also has some possible decrease in the p_1-surplus coming from Step (b), which is described by (7.24):

$$\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) = \nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right) + N(A_{p_1} - B_{p_1}).$$

In particular, we have from (7.15), (7.16) and the triangle inequality that

$$\frac{1}{N}\left\|\nu_2\left(\prod \mathcal{B}^{(1)}\right), \nu_3\left(\prod \mathcal{B}^{(1)}\right)\right\|_\gamma \leq \alpha_1 + \alpha_2. \tag{7.25}$$

Each element removed in Step (a) reduces the t-excess, while each element $p\lceil t/p\rceil$ added in Step (b) increases the t-excess by $\log\frac{\lceil t/p\rceil}{t/p}$, so each large prime $t/K < p \leq N$ contributes a net of $\lfloor \frac{N}{p}\rfloor \log\frac{\lceil t/p\rceil}{t/p} = f_{N/t}(p/N)$ to the t-excess. Thus by (7.6), (7.7) we have

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(1)}) \leq \delta_1 + \delta_2. \tag{7.26}$$

Now we bring the multiset $\mathcal{B}^{(1)}$ into balance at small and medium primes $3 < p \leq t/K$. We make the following observations:

(C) If an element in $\mathcal{B}^{(1)}$ is divisible by some small or medium prime $3 < p \leq t/K$, and one replaces p by $\lceil p\rceil^{\langle 2,3\rangle}_{4.5}$ in the factorization of that element, then the p-deficit decreases by one, while (by Lemma 2.1) the t-excess increases by at most $\kappa_{4.5}$, and the quantity $\|(\nu_2(\prod \mathcal{B}^{(1)}), \nu_3(\prod \mathcal{B}^{(1)}))\|_\gamma$ increases by at most $\frac{2}{\log 12}(\log p + \kappa_{**})$. All other p_1-surpluses or p_1-deficits for $p_1 \neq 2, 3, p$ remain unaffected.

(D) If one adds an element of the form $m\lceil t/m\rceil_{4.5}^{\langle 2,3\rangle}$ to $\mathcal{B}^{(1)}$ for some $m \leq t/K$ that is the product of small or medium primes $3 < p \leq t/K$, then the p-surpluses at small or medium primes p decrease by $\nu_p(m)$, while (by Lemma 2.1) the t-excess increases by at most $\kappa_{4.5}$, and the quantity $\sup_{p_0=2,3} \frac{p_0-1}{N}\nu_{p_0}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right)$ increases by at most $\frac{2}{\log 12}(\log(t/m)+\kappa_{**})$. The p-surpluses or p-deficits at medium or large primes remain unaffected.

With these observations in mind, we perform the following modifications to the multiset $\mathcal{B}^{(1)}$.

(c) If there is a p_1-deficit $\nu_{p_1}(\prod \mathcal{B}^{(1)}/N!) > 0$ at some small or medium prime $3 < p_1 \leq t/K$, then we perform the replacement of p_1 in one of the elements of $\mathcal{B}^{(1)}$ with $\lceil p_1\rceil_{4.5}^{\langle 2,3\rangle}$ as per observation (C), repeated $\nu_{p_1}(\prod \mathcal{B}^{(1)}/N!)$ times, in order to eliminate all such deficits.

(d) If there is a p-surplus $\nu_p(\prod N!/\mathcal{B}^{(1)}) > 0$ at some medium prime $K < p \leq t/K$, we add the element $p\lceil t/p\rceil_{4.5}^{\langle 2,3\rangle}$ to $\mathcal{B}^{(1)}$ as per observation (D), $\nu_p(\prod N!/\mathcal{B}^{(1)})$ times, in order to eliminate all such surpluses at medium primes.

(d') If there are p-surpluses $\nu_p(\prod N!/\mathcal{B}^{(1)}) > 0$ at some small primes $3 < p \leq K$, we multiply all these primes together, then apply the greedy algorithm to factor them into products m in the range $t/K^2 < m \leq t/K$, plus at most one exceptional product in the range $1 < m \leq t/K$. For each of these m, add $m\lceil t/m\rceil_{4.5}^{\langle 2,3\rangle}$ to $\mathcal{B}^{(1)}$ as per observation (D), to eliminate all such surpluses at small primes.

Call the multiset formed from $\mathcal{B}^{(1)}$ formed as the outcome of applying Steps (c), (d), (d') as $\mathcal{B}^{(2)}$. The product of all the primes arising in Step (d') has logarithm equal to

$$\sum_{3<p_1\leq K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right)\right|_{\log p_1,0} = \sum_{3<p_1\leq K}\left|\nu_{p}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)\right|_{\log p_1,0}$$

and hence the number of non-exceptional m arising in (d') is at most

$$\sum_{3<p_1\leq K}\left|\nu_{p}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)\right|_{\frac{\log p_1}{\log(t/K^2)},0}.$$

The total excess of $\mathcal{B}^{(2)}$ is increased in Step (c) by at most

$$\kappa_{4.5}\sum_{3<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right)\right|_{0,1} = \kappa_{4.5}\sum_{3<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)+N(A_{p_1}-B_{p_1})\right|_{0,1},$$

in Step (d) by at most

$$\kappa_{4.5}\sum_{K<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right)\right|_{1,0} = \kappa_{4.5}\sum_{K<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)+NA_{p_1}\right|_{1,0},$$

and in Step (c) by at most

$$\kappa_{4.5}\left(1+\sum_{3<p_1\leq K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)\right|_{\frac{\log p_1}{\log(t/K^2)},0}\right).$$

From the triangle inequality and (7.26), (7.8), (7.9), (7.10), (7.11), we then have

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(2)}) \leq \sum_{i=1}^{6} \delta_i. \tag{7.27}$$

Similarly, the quantity $\frac{1}{N}\|(\nu_2(\prod \mathcal{B}^{(1)}), \nu_3(\prod \mathcal{B}^{(1)}))\|_\gamma$ is increased in Step (c) by at most

$$\frac{2}{N\log 12}\sum_{3<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)+N(A_{p_1}-B_{p_1})\right|_{0,\log p_1+\kappa_{**}},$$

in Step (d) by at most

$$\frac{2}{N\log 12}\sum_{K<p_1\leq t/K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)+NA_{p_1}\right|_{\log(t/p_1)+\kappa_{**},0},$$

and in Step (d') by at most the sum of

$$\frac{2}{N\log 12}\sum_{3<p_1\leq K}\left|\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(0)}}\right)+N(A_{p_1}-B_{p_1})\right|_{\log(K^2)+\kappa_{**},0}$$

and

$$\frac{2}{N\log 12}\left(\log t+\kappa_{**}\right)$$

so by (7.25), (7.17), (7.18), (7.19), (7.20), and the triangle inequality we have

$$\frac{1}{N}\|(\nu_2(\prod \mathcal{B}^{(2)}), \nu_3(\prod \mathcal{B}^{(3)}))\|_\gamma \leq \sum_{i=1}^{6} \alpha_i. \tag{7.28}$$

By construction, the tuple $\mathcal{B}^{(2)}$ is t-admissible, and in balance at all small, medium, and large primes $p > 3$; thus $N!/\prod \mathcal{B}^{(2)} = 2^n 3^m$ for some integers n, m. From (7.28), (7.5), (2.10), (7.21) we have

$$\begin{aligned}
n - 2\gamma_2 m &= \nu_2(N!) - 2\gamma_2\nu_3(N!) - \left(\nu_2\left(\prod \mathcal{B}^{(2)}\right) - 2\gamma_2\nu_3\left(\prod \mathcal{B}^{(2)}\right)\right)\\
&\geq \nu_2(N!) - 2\gamma_2\nu_3(N!) - N(1-\gamma_2)\sum_{i=1}^{6}\alpha_i\\
&> N - \frac{\log N}{\log 2} - 1 - \gamma_2 N - N(1-\gamma_2)(1-\alpha_7)\\
&= N(1-\gamma_2)\alpha_7 - \frac{\log(2N)}{\log 2}\\
&\geq 0
\end{aligned}$$

and similarly

$$\begin{aligned}
2m - \gamma_3 n &= 2\nu_3(N!) - \gamma_3\nu_2(N!) - \left(2\nu_3\left(\prod \mathcal{B}^{(2)}\right) - \gamma_3\nu_2\left(\prod \mathcal{B}^{(2)}\right)\right)\\
&\geq 2\nu_3(N!) - \gamma_3\nu_2(N!) - N(1-\gamma_3)\sum_{i=1}^{6}\alpha_i\\
&> N - 2\frac{\log N}{\log 3} - 2 - \gamma_3 N - N(1-\gamma_3)(1-\alpha_7)\\
&= N(1-\gamma_3)\alpha_7 - 2\frac{\log(3N)}{\log 3}\\
&\geq 0.
\end{aligned}$$

From (7.3) and Cramer's rule we conclude that that $(n, 2m)$ lies in the non-negative linear span of $(n_2, 2m_2)$, $(n_3, 2m_3)$, thus

$$(n, 2m) = \beta_2(n_2, 2m_2) + \beta_3(n_3, 2m_3) \tag{7.29}$$

for some reals $\beta_2, \beta_3 \geq 0$. We now create the multiset $\mathcal{B}^{(3)}$ by adding $\lfloor\beta_2\rfloor$ copies of $2^{n_2}3^{m_2}$ and $\lfloor\beta_3\rfloor$ copies of $2^{n_3}3^{m_3}$ to $\mathcal{B}^{(2)}$. By (7.2), this multiset remains t-admissible, and each element added increases the t-excess by at most κ_*. The number of such elements can be upper bounded using (7.29), (2.10) as

$$\begin{aligned}
\lfloor\beta_2\rfloor + \lfloor\beta_3\rfloor &\leq \beta_2 + \beta_3\\
&\leq \frac{1}{\log t}\left(\beta_2(n_2\log 2 + m_2\log 3) + \beta_3(n_3\log 2 + m_3\log 3)\right)\\
&= \frac{1}{\log t}(n\log 2 + m\log 3)\\
&\leq \frac{1}{\log t}((\nu_2(N!) - NB_2)\log 2 + (\nu_3(N!) - NB_3)\log 3)\\
&\leq \frac{1}{\log t}\left(N\log 2 + \frac{N}{2}\log 3 - NB_2\log 2 - NB_3\log 3\right)\\
&= \frac{N\log 12}{2\log t} - \frac{N(B_2\log 2 + B_3\log 3)}{\log t}.
\end{aligned}$$

By (7.27), (7.12), we thus have

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \sum_{i=1}^{7}\delta_i. \tag{7.30}$$

Meanwhile by construction we see that $\mathcal{B}^{(3)}$ is a subfactorization of $N!$ that is in balance at all non-tiny primes, with tiny prime surpluses bounded by

$$\nu_2\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right) \leq n_2 + n_3; \quad \nu_3\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right) \leq m_2 + m_3.$$

and thus by (7.2), (7.13), we thus have

$$\frac{1}{N}\sum_p \nu_p\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right)\log p \leq \frac{\log 2^{n_2}3^{m_2} + \log 2^{n_3}3^{m_3}}{N} \leq \delta_8$$

and thus by (7.30), (7.4) we have

$$\mathcal{E}_t(\mathcal{B}^{(3)}) + \sum_p v_p \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) \log p \leq \log N! - N \log t.$$

Applying Lemma 5.2, we conclude that $t(N) \geq t$ as claimed.

8. Estimating terms

In order to use Proposition 7.1 for a given choice of N, t, we need to find a t-admissible tuple $\mathcal{B}^{(0)}$ and parameters $K, \kappa_*, \gamma_2, \gamma_3$ obeying (7.1) as well as good upper bounds on the quantities δ_i, $i = 1, \dots, 8$ and α_i, $i = 1, \dots, 7$, that can either be evaluated asymptotically or numerically. Many of the terms here will be straightforward to estimate; we discuss only the more difficult ones.

We introduce a further natural number parameter A and define

$$\sigma := \frac{3N}{At}. \tag{8.1}$$

We let $\mathcal{B}^{(0)}$ be the multiset of 3-rough elements of the interval $(t, t(1+\sigma)]$, with each element repeated precisely A times. This is clearly t-admissible. It has no presence at tiny primes, so

$$\alpha_1 = 0. \tag{8.2}$$

We use the notation $\sum^*$ to denote summation restricted to 3-rough numbers, thus for instance $\sum^*_{a<k\leq b} 1$ denotes the number of 3-rough numbers in $(a, b]$. We have a simple estimate for such counts:

Lemma 8.1. *For any interval $(a, b]$ with $0 \leq a \leq b$ one has $\sum^*_{a<k\leq b} 1 = \frac{b-a}{3} + O_{\leq}(4/3)$.*

Proof. By the triangle inequality, it suffices to show that $\sum^*_{0<k\leq x} 1 - \frac{x}{3} = O_{\leq}(2/3)$ for all $x \geq 0$. The claim is easily verified for $0 \leq x \leq 6$, and the left-hand side is 6-periodic in x, giving the claim; see Figure 8. $\square$

This lets us estimate δ_1:

Lemma 8.2. *We have*

$$\delta_1 \leq \frac{3N}{2tA} + \frac{4}{N}.$$

Proof. By definition, we have

$$\mathcal{E}_t(\mathcal{B}^{(1)}) = A \sum^*_{t<n\leq t(1+\sigma)} \log \frac{n}{t}.$$

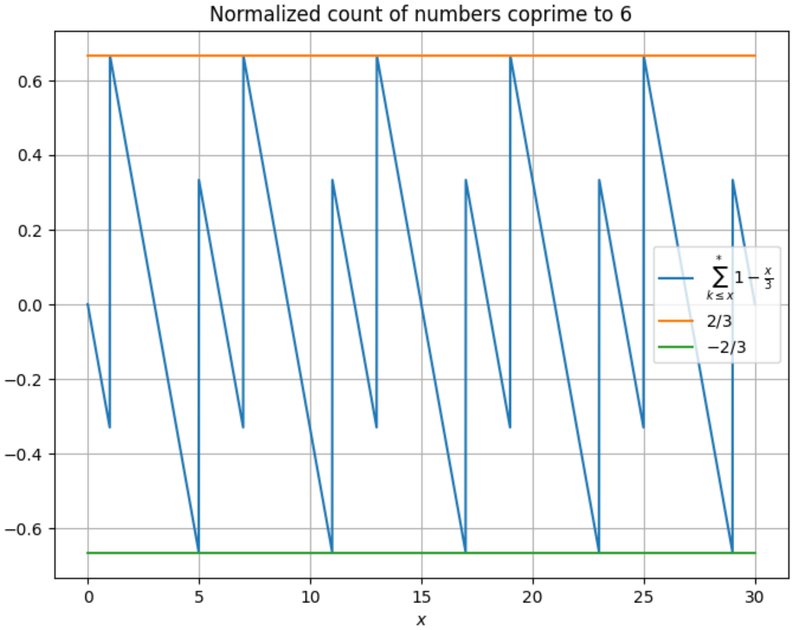

FIGURE 8. The function $\sum^*_{k\le x} 1 - \frac{x}{3}$.

By the fundamental theorem of calculus, this is

$$A\int_0^{t\sigma} \sum^*_{t<n\le t+h} 1 \,\frac{dh}{t+h}.$$

Bounding $\frac{1}{t+h}$ by $\frac{1}{t}$ and applying Lemma 8.1, (8.1), we conclude that

$$\mathcal{E}_t(\mathcal{B}^{(1)}) \le A\int_0^{3N/A} \left(\frac{h}{3}+\frac{4}{3}\right)\frac{dh}{t} = \frac{3N^2}{2tA} + 4.$$

and the claim follows. □

To construct $\gamma_2, \gamma_3, \kappa_*, n_2, m_2, n_3, m_3$, we introduce another parameter $L \ge 1$ and assume that

$$t > 3L. \tag{8.3}$$

We define n_2, n_3, m_2, n_3 by setting

$$2^{n_2}3^{m_2} := 2^{n_0}\lceil t/2^{n_0}\rceil^{\langle 2,3\rangle}; \quad 2^{n_3}3^{m_3} := 3^{m_0}\lceil t/3^{m_0}\rceil^{\langle 2,3\rangle}$$

where $2^{n_0}, 3^{m_0}$ are the largest powers of 2, 3 respectively that are at most t/L. By construction and (2.2), (7.2) holds with

$$\kappa_* = \kappa_L. \tag{8.4}$$

We have

$$2m_2 \le \frac{2}{\log 3}\log\lceil t/2^{n_0}\rceil^{\langle 2,3\rangle} \le \frac{2}{\log 3}(\log(2L)+\kappa_L)$$

and

$$n_2 \ge n_0 \ge \frac{\log t - \log(2L)}{\log 2};$$

similarly

$$n_3 \leq \frac{1}{\log 2}(\log(3L) + \kappa_L)$$

and

$$2m_3 \geq \frac{2(\log t - \log(3L))}{\log 3}.$$

We conclude that (7.3) holds with

$$\begin{aligned} \gamma_2 &:= \frac{2\log 2}{\log 3}\frac{\log(2L)+\kappa_L}{\log t - \log(2L)} \\ \gamma_3 &:= \frac{\log 3}{2\log 2}\frac{\log(3L)+\kappa_L}{\log t - \log(3L)}; \end{aligned} \tag{8.5}$$

one can of course also take larger values of γ_2, γ_3 if desired. This lets us compute the quantity κ_{**} defined in (7.22).

To estimate δ_3, α_3 we use

Lemma 8.3. *For every $3 < p \leq t/K$, one has*

$$\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) = O_{\leq}\left(\frac{4A+3}{3}\left\lceil\frac{\log N}{\log p}\right\rceil\right). \tag{8.6}$$

Proof. One has

$$\begin{aligned} \nu_p(\prod \mathcal{B}^{(1)}) &= A \sideset{}{^*}\sum_{t<n\leq t(1+\sigma)} \nu_p(n) \\ &= A \sum_{1\leq j\leq \frac{\log N}{\log p}} \sideset{}{^*}\sum_{t/p^j<n\leq t(1+\sigma)/p^j} 1 \\ &= A \sum_{1\leq j\leq \frac{\log N}{\log p}} \left(\frac{N}{p^j A} + O_{\leq}(4/3)\right) \\ &= \frac{N}{p-1} - O^+_{\leq}\left(\frac{1}{p-1}\right) + O_{\leq}\left(\frac{4A}{3}\left\lceil\frac{\log N}{\log p}\right\rceil\right) \\ &= \frac{N}{p-1} - O^+_{\leq}\left(\left\lceil\frac{\log N}{\log p}\right\rceil\right) + O_{\leq}\left(\frac{4A}{3}\left\lceil\frac{\log N}{\log p}\right\rceil\right). \end{aligned}$$

Meanwhile, from (2.10) one has

$$\nu_p(N!) = \frac{N}{p-1} - O^+_{\leq}\left(\left\lceil\frac{\log N}{\log p}\right\rceil\right)$$

and the claim follows. ☐

Corollary 8.4. *One has*

$$\delta_3 \leq \frac{(4A+3)\kappa_{4.5}}{3N}\left(\pi(t/K) + \frac{\log N}{\log 5}\pi(\sqrt{N})\right)$$

and

$$\alpha_3 \leq \frac{2(4A+3)}{3N\log 12}\left(\log\frac{t}{K}+\kappa_{**}\right)\left(\pi(t/K)+\frac{\log N}{\log 5}\pi(\sqrt{N})\right).$$

Proof. This is immediate from Lemma 8.3 and (7.17), (7.8) after noting that $\lfloor\frac{\log N}{\log p}\rfloor \leq 1+\frac{\log N}{\log 5}1_{p\leq\sqrt{N}}$ for $3 < p \leq t/K$. □

The main quantities left to estimate are the quantities $\delta_4, \delta_5, \alpha_4, \alpha_5$ that involve A_{p_1}. By construction of $\mathcal{B}^{(0)}$, we have

$$A_{p_1} = \frac{1}{N}\sum_m^* \nu_{p_1}(m)\sum_{\frac{t}{K},\frac{t}{m}<p\leq\frac{t(1+\sigma)}{m}} A.$$

In particular, for $p > K(1+\sigma)$ the quantity A_{p_1} vanishes entirely:

$$A_{p_1} = 0. \tag{8.7}$$

For the remaining primes $3 < p \leq K(1+\sigma)$ one has

$$A_{p_1} = \frac{A}{N}\sum_{m\leq K(1+\sigma)}^* \nu_{p_1}(m)\left(\pi\left(\frac{t(1+\sigma)}{m}\right)-\pi\left(\frac{t}{\min(m,K)}\right)\right). \tag{8.8}$$

In practice, these expressions can be adequately controlled by Lemma 2.2, as can the quantities B_{p_1}.

9. The asymptotic regime

With the above estimates, we can now establish the lower bound in Theorem 1.3(iv). Thus we aim to show that $t(N) \geq t$ for sufficiently large N, where

$$t := \frac{N}{e} - \frac{c_0 N}{\log N} + \frac{N}{\log^{1+c_1} N} \asymp N \tag{9.1}$$

and $0 < c_1 < 1$ is a small absolute constant. We use the construction of the previous section with the parameters

$$A := \lfloor \log^2 N \rfloor \tag{9.2}$$

$$K := \lfloor \log^3 N \rfloor \tag{9.3}$$

$$L := N^{0.1}, \tag{9.4}$$

so from (8.1) one has

$$\sigma = \frac{3N}{tA} \asymp \frac{1}{A} \asymp \frac{1}{\log^2 N}. \tag{9.5}$$

The conditions (7.1), (8.3) are easily verified for N large enough.

By (8.4), (9.4), and Lemma 2.1(ii) we have

$$\kappa_* \ll \log^{-c} N$$

for some absolute constant $c > 0$. From (8.5), (9.1), (9.4) we have

$$\gamma_2 = \frac{1}{10}\frac{2\log 2}{\log 3} + O\left(\frac{1}{\log N}\right), \quad \gamma_3 = \frac{1}{10}\frac{\log 3}{2\log 2} + O\left(\frac{1}{\log N}\right)$$

and hence by (7.22), (2.6), (2.7)

$$\kappa_{**} \ll 1.$$

By Proposition 7.1, it thus suffices to establish the inequalities (7.4), (7.5). Several of the quantities $\delta, \delta_i, \alpha_i$ can now be immediately estimated using (8.2), (8.2), Corollary 8.4, (2.11), and the prime number theorem:

$$\begin{aligned}
\delta_1 &\ll \frac{1}{\log^2 N} \\
\delta_3 &\ll \frac{1}{\log^2 N} \\
\delta_6 &\ll \frac{1}{N} \\
\delta_7 &\ll \frac{1}{\log^{1+c} N} \\
\delta_8 &\ll \frac{\log N}{N} \\
\delta &= \frac{ec_0}{\log N} + \frac{e}{\log^{1+c_1} N} + O\left(\frac{1}{\log^2 N}\right)
\end{aligned}$$

$$\begin{aligned}
\alpha_1 &= 0 \\
\alpha_3 &\ll \frac{1}{\log N} \\
\alpha_6 &\ll \frac{\log N}{N} \\
\alpha_7 &\ll \frac{\log N}{N}.
\end{aligned}$$

On the interval $(t/NK, 1]$, the function $f_{N/t}$ is piecewise monotone with $O(K)$ pieces, and bounded by 1, so its augmented total variation norm is $O(K)$. Applying (7.7) and Lemma 2.2 (with classical error term), we have

$$\begin{aligned}
\delta_2 &\leq \frac{1}{\log(t/K)} \int_{t/NK}^{1} f_{N/t}(x)\, dx + O\left(\frac{1}{\log^2 N}\right) \\
&\leq \frac{1}{\log N} \int_{1/eK}^{N/et} f_{N/t}(etx/N)\, dx + O\left(\frac{1}{\log^2 N}\right)
\end{aligned}$$

where we have used (1.8) to manage error terms. As in the proof of Proposition 5.4, the function $f_{N/t}(etx/N)$ differs from $f_e(x)$ outside of an exceptional set of measure $O(1/\log N)$, and hence by (1.6) (and (1.8)) we have

$$\delta_2 \leq \frac{ec_0}{\log N} + O\left(\frac{1}{\log^2 N}\right).$$

To finish the verification of the conditions (7.4), (7.5), it will suffice to show that

$$\delta_4, \delta_5 \ll \frac{(\log\log N)^{O(1)}}{\log^2 N} \tag{9.6}$$

and

$$\alpha_2, \alpha_4, \alpha_5 \ll \frac{(\log\log N)^{O(1)}}{\log N}. \tag{9.7}$$

By Mertens' theorem (or Lemma 2.2) and (7.9), (7.10), (7.16), (7.18), (7.19), (9.5), it suffices to show that

$$A_{p_1}, B_{p_1} \ll \frac{(\log\log N)^{O(1)}}{p_1 \log N} \tag{9.8}$$

for all $p_1 \leq K(1+\sigma)$ (recalling from (8.7) that A_{p_1} vanishes for any larger p_1), as well as the variant

$$|A_{p_1} - B_{p_1}|_{0,1} \ll \frac{(\log\log N)^{O(1)}}{p_1 \log^2 N} \tag{9.9}$$

for $3 < p_1 \leq K$.

For (9.8) we use (8.8), (7.24), and the crude bound

$$\nu_{p_1}(m) \ll 1_{p_1|m} \log\log N \tag{9.10}$$

for $m \leq K(1+\sigma)$, and reduce to showing that

$$\frac{A}{N} \sum_{m \leq K(1+\sigma)} 1_{p_1|m} \left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{\min(m,K)}\right)\right). \ll \frac{(\log\log N)^{O(1)}}{p_1 \log N}$$

and

$$\frac{1}{N} \sum_{m \leq K} 1_{p_1|m} \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \left\lfloor \frac{N}{p} \right\rfloor \ll \frac{(\log\log N)^{O(1)}}{p_1 \log N}.$$

But from the Brun–Titchmarsh inequality (or Lemma 2.2) and (9.5) one has

$$\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{\min(m,K)}\right) \ll \frac{t\sigma}{m \log N} \ll \frac{N}{Am \log N}$$

and

$$\sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \left\lfloor \frac{N}{p} \right\rfloor \ll \frac{tm}{m^2 \log N} \ll \frac{N}{m \log N}$$

and the claim then follows from summing the harmonic series.

It remains to show (9.9). For $3 < p_1 \leq K$, we see from (8.8), (9.5), (9.10) and Lemma 2.2 (with classical error term) that

$$\begin{aligned} A_{p_1} &\geq \frac{1}{N} \sum_{m \leq K(1+\sigma)}^{*} \nu_{p_1}(m) \left(\frac{At\sigma}{m \log N} + O\left(\frac{(\log\log N)^{O(1)} At\sigma}{m \log^2 N} \right) \right) \\ &= \frac{1}{\log N} \sum_{m \leq K(1+\sigma)}^{*} \nu_{p_1}(m) \frac{3}{m} + O\left(\frac{(\log\log N)^{O(1)}}{\log^2 N} \right) \\ &= \frac{1}{\log N} \sum_{m \leq K}^{*} \nu_{p_1}(m) \frac{3}{m} + O\left(\frac{(\log\log N)^{O(1)}}{\log^2 N} \right) \end{aligned}$$

and similarly from (7.24), (9.10), and Lemma 2.2 (again with classical error term)

$$\begin{aligned} B_{p_1} &\leq \frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \frac{N}{p} \\ &\leq \frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \left(\frac{N}{\log(t/m)} \int_{t/m}^{t/(m-1)} \frac{dx}{x} + O\left(\frac{N}{\log^{10} N} \right) \right) \\ &\leq \frac{1}{\log N} \sum_{m \leq K} \nu_{p_1}(m) \log \frac{m}{m-1} + O\left(\frac{(\log\log N)^{O(1)}}{\log^2 N} \right) \end{aligned}$$

so it will suffice to establish the inequality

$$\sum_{m \leq K} \nu_{p_1}(m) \log \frac{m}{m-1} \leq \sum_{m \leq K}^{*} \nu_{p_1}(m) \frac{3}{m} \tag{9.11}$$

for all $p_1 > 3$.

Writing $\nu_{p_1}(m) = \sum_{j \geq 1} 1_{p_1^j | m}$, it suffices to show that

$$\sum_{m \leq K; p^j | m} \frac{3}{m} 1_{(m,6)=1} - \log \frac{m}{m-1} \geq 0.$$

Making the change of variables $m = p_1^j n$, it suffices to show that

$$\sum_{n \leq K'} \frac{3}{n} 1_{(n,6)=1} - p_1^j \log \frac{p_1^j n}{p_1^j n - 1} \geq 0$$

for any $K' > 0$. Using the bound

$$\log \frac{p_1^j n}{p_1^j n - 1} = \int_{p_1^j n - 1}^{p_1^j n} \frac{dx}{x} \leq \frac{1}{p_1^j n - 1}$$

and $p^j \geq 5$, we have

$$p_1^j \log \frac{p_1^j n}{p_1^j n - 1} \leq \frac{1}{n - 0.2}$$

and so it suffices to show that

$$\sum_{n \leq K'}^{*} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n - 0.2} \geq 0. \tag{9.12}$$

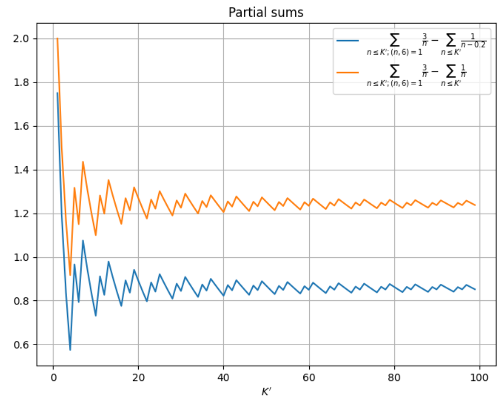

FIGURE 9. A plot of (9.12), (9.13).

Since

$$\sum_{n=1}^{\infty} \frac{1}{n-0.2} - \frac{1}{n} = \psi(0.8) - \psi(1) = 0.353473\dots,$$

where ψ here denotes the digamma function rather than the von Mangoldt summatory function, it will suffice to show that

$$\sum_{n\leq K'} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n} \geq 0.4. \tag{9.13}$$

This can be numerically verified for $K' \leq 100$, with substantial room to spare for K' large; see Figure 9. On a block $6a-1 \leq n \leq 6a+4$ with $a > 1$, the sum is positive:

$$\begin{aligned}\sum_{6a-1\leq n\leq 6a+4}^{*} \frac{3}{n} - \frac{1}{n} &= \left(\frac{1}{6a-1} - \frac{1}{6a}\right) + \left(\frac{1}{6a-1} - \frac{1}{6a+2}\right) \\ &\quad + \left(\frac{1}{6a+1} - \frac{1}{6a+3}\right) + \left(\frac{1}{6a+1} - \frac{1}{6a+4}\right) \\ &> 0.\end{aligned}$$

The inequality for $K' > 100$ is then easily verified from the $K' \leq 100$ data and the triangle inequality.

10. GUY–SELFRIDGE CONJECTURE

We now establish the Guy–Selfridge conjecture $t(N) \geq N/3$ in the range

$$N \geq N_0 := 10^{11}.$$

We will apply Proposition 7.1 with the construction in Section 8 and the choice of parameters

$$\begin{aligned} t &:= N/3 \\ A &:= 190 \\ K &:= 252 \\ L &:= 4.5; \end{aligned}$$

the choice of A and K was obtained after some numerical experimentation. In particular, by (8.1) we have

$$\sigma = \frac{3N}{At} = \frac{9}{190} = 0.047368\dots$$

and so

$$K(1+\sigma) = 263.936\dots$$

One can readily check the required conditions (7.1), (8.3) for $N \geq N_0$, so it remains to verify the hypotheses (7.4), (7.5) of Proposition 7.1 in this range. Some of the quantities in these hypotheses involve sums over large ranges, such as $(t/K, N]$; but one can use Lemma 2.2 to obtain adequate upper or lower bounds on such quantities, leaving one with sums over short ranges such as $p \leq K$ or $p \leq K(1+\sigma)$. Our bounds are designed to be monotone non-increasing in N, so the task reduces to evaluating these short sums for $N = N_0$, which can be accomplished with simple computer code, which can be found at `https://github.com/teorth/erdos-guy-selfridge/blob/main/src/python/calculations.py`. As it turns out, we will have a tiny bit of room to spare, in that we can show

$$\sum_{i=1}^{8} \delta_i \leq 0.9729\delta \tag{10.1}$$

and

$$\sum_{i=1}^{7} \alpha_i \leq 0.9697. \tag{10.2}$$

We now bound some of the terms appearing in the above expression. From Lemma 2.1 we have

$$\kappa_{4.5} = \log\frac{4}{3} = 0.28768\dots.$$

From (8.5) one can take

$$\gamma_2 := \frac{2\log 2}{\log 3}\frac{\log(2L)+\kappa_L}{\log(N_0/3)-\log(2L)} = 0.1423165\dots$$

and

$$\gamma_3 := \frac{\log 3}{2\log 2}\frac{\log(3L)+\kappa_L}{\log(N_0/3)-\log(3L)} = 0.1059116\dots$$

and so by (7.22) and some calculation

$$\kappa_{**} \leq 6.830101\dots.$$

From (2.11) one has

$$\delta \geq \log N - \log t = \log\frac{3}{e} = 0.0986122\dots.$$

From (8.2) on has

$$\delta_1 \leq \frac{9}{2A} + \frac{4}{N_0} \leq 0.240176\delta.$$

From (7.7) and Lemma 2.2 one has

$$\begin{aligned}\delta_2 &\leq \frac{1}{\log(t/K)}\int_{1/3K}^{1} f_3(x)\,dx + \frac{1}{\log(t/K)}\|f_3\|_{\mathrm{TV}((1/3K,1])}\frac{E(N)}{N}\\ &\leq \frac{1}{\log(N_0/3K)}0.919507 + \frac{1}{\log(N_0/3K)}996.196\frac{E(N_0)}{N_0}\\ &\leq 0.50025\delta.\end{aligned}$$

From Corollary 8.4 and (2.13) one has

$$\begin{aligned}\delta_3 &\leq \frac{(4A+3)\kappa_{4.5}}{3}\left(\frac{1}{3K\log(N/3K)} + \frac{1.2762}{3K\log^2(N/3K)} + \frac{\log N}{\log 5\sqrt{N}\log\sqrt{N}} + \frac{1.2762\log N}{\log 5\sqrt{N}\log^2\sqrt{N}}\right)\\ &\leq \frac{(4A+3)\kappa_{4.5}}{3}\left(\frac{1}{3K\log(N_0/3K)} + \frac{1.2762}{3K\log^2(N_0/3K)} + \frac{\log N_0}{\log 5\sqrt{N_0}\log\sqrt{N_0}} + \frac{1.2762\log N_0}{\log 5\sqrt{N_0}\log^2\sqrt{N_0}}\right)\\ &\leq 0.059274\delta.\end{aligned}$$

We skip $\delta_4, \delta_5, \delta_7$ for now. From (7.11) we have

$$\delta_6 \leq \frac{\kappa_{4.5}}{N_0} \leq 3\times 10^{-11}\delta$$

and from (7.13) we have

$$\delta_8 \leq \frac{2(\log(N_0/3)+\kappa_{4.5})}{N_0} \leq 6\times 10^{-10}\delta$$

so these two terms are negligible in the analysis.

From (8.2) we have

$$\alpha_1 = 0.$$

We skip $\alpha_2, \alpha_4, \alpha_5$ for now. From Corollary 8.4 and (2.13) one has

$$\begin{aligned}\alpha_3 &\leq \frac{2(4A+3)}{3\log 12}\left(\log\frac{N}{3K} + \kappa_{**}\right)\\ &\quad\times\left(\frac{1}{3K\log(N/3K)} + \frac{1.2762}{3K\log^2(N/3K)} + \frac{\log N}{\log 5\sqrt{N}\log\sqrt{N}} + \frac{1.2762\log N}{\log 5\sqrt{N}\log^2\sqrt{N}}\right).\end{aligned}$$

Expanding out the product, one can check that all terms are non-increasing in N; so we may substitute N_0 for N in the right-hand side, which after some calculation gives

$$\alpha_3 \leq 0.417501.$$

From (7.20) we have

$$\begin{aligned}\alpha_6 &\leq \frac{2}{\log 12}\left(\frac{\log(N_0/3)}{N_0} + \frac{\kappa_{**}}{N_0}\right)\\ &\leq 3\times 10^{-10}\end{aligned}$$

and similarly from (7.21) we have

$$\begin{aligned}\alpha_7 &\leq \max\left(\frac{\log(2N_0)}{(1-\gamma_2)N_0\log 2}, \frac{2\log(3N_0)}{(1-\gamma_3)N_0\log 3}\right)\\ &\leq 6\times 10^{-10}\end{aligned}$$

so the contribution of these two terms are negligible.

The remaining terms $\delta_4, \delta_5, \delta_7, \alpha_2, \alpha_4, \alpha_5$ to estimate involve the quantities A_{p_1}, B_{p_1} defined in (7.23), (7.24). For B_{p_1}, we can split it as

$$B_{p_1} = \sum_{m\leq K} \nu_{p_1}(m) \sum_{k: a_{k,m}<b_{k,m}} k\frac{1}{N}(\pi(Nb_{k,m}) - \pi(Na_{k,m}))$$

where

$$a_{k,m} := \max\left(\frac{1}{3m}-, \frac{1}{k}\right); \quad b_{k,m} := \max\left(\frac{1}{3(m-1)}-, \frac{1}{k-1}\right)$$

where the $-$ denotes the subtraction of an infinitesimal quantity to reflect the restriction to the range $\frac{t}{m} \leq p < \frac{t}{m-1}$ rather than $\frac{t}{m} < p \leq \frac{t}{m-1}$. Using Lemma 2.2, we can upper bound this quantity by

$$B_{p_1} \leq \sum_{m\leq K} \nu_{p_1}(m) \sum_{k: a_{k,m}<b_{k,m}} k\frac{1}{\log(Na_{k,m})}\left(a_{k,m} - b_{k,m} + 2\frac{E(N_0 b_{k,m})}{N_0}\right)$$

and lower bound it by

$$B_{p_1} \geq \sum_{m\leq K} \nu_{p_1}(m) \sum_{k: a_{k,m}<b_{k,m}} k\frac{1}{\log(Nb_{k,m})}\left(\left(1-\frac{2}{\sqrt{a_{k,m}}}\right)(a_{k,m}-b_{k,m}) + 2\frac{E(N_0 b_{k,m})}{N_0 b_{k,m}}\right).$$

After some calculation, these bounds inserted into (7.16) give

$$\alpha_2 \leq 0.260087$$

and in (7.12) they give

$$\delta_7 \leq 0.112941\delta.$$

As for the A_{p_1}, we know from (8.7) that this vanishes unless $3 < p_1 \leq K(1+\sigma)$. From (8.8) and Lemma 2.2 one has the upper bound

$$A_{p_1} \leq A \sideset{}{^*}\sum_{m\leq K(1+\sigma)} \nu_{p_1}(m)\frac{1}{\log(N/3\min(m,K))}\left(\frac{1+\sigma}{3m} - \frac{1}{3\min(m,K)} + \frac{2E(N_0(1+\sigma)/3m)}{N_0}\right)$$

and the lower bound

$$A_{p_1} \geq A \sideset{}{^*}\sum_{m\leq K(1+\sigma)} \nu_{p_1}(m)\frac{1}{\log(N(1+\sigma)/3m)}\left(\left(1-\frac{2}{\sqrt{N_0(1+3\sigma)/3m}}\right)\left(\frac{1+\sigma}{3m} - \frac{1}{3\min(m,K)}\right) - \frac{2E(N_0(1}{N}\right.$$

Using these bounds, one can verify by direct computation that $A_{p_1} \geq B_{p_1}$ for all $3 < p_1 \leq K(1+\sigma)$ (cf. (9.11)), and from (7.9), (7.10), (7.18), (7.19) one can then verify that

$$\begin{aligned}\delta_4 &\leq 0.001212\delta\\ \delta_5 &\leq 0.057447\delta\\ \alpha_4 &\leq 0.008523\\ \alpha_5 &\leq 0.283524\end{aligned}$$

and the claims (10.1), (10.2) follow by summing all the bounds.

Appendix A. Distance to the next 3-smooth number

We now establish the various claims in Lemma 2.1. We begin with part (iii). The claim (2.3) is immediate from (2.2), (2.1). Now prove (2.4), (2.5). If we write $\lceil x/12^a\rceil^{\langle 2,3\rangle} = 2^b 3^c$, then by (2.2) we have

$$b\log 2 + c\log 3 \leq \log x - a\log 12 + \kappa_L,$$

while from definition of a we have

$$\log x - a\log 12 \leq \log(12L). \tag{A.1}$$

We now compute

$$\begin{aligned}\frac{\nu_2(\lceil x\rceil_L^{\langle 2,3\rangle}) - 2\gamma\nu_3(\lceil x\rceil_L^{\langle 2,3\rangle})}{1-\gamma} &= \frac{2a+b-2\gamma(a+c)}{1-\gamma}\\ &\leq 2a + \frac{\log x - a\log 12 + \kappa_L}{(1-\gamma)\log 2}\\ &= \frac{2\log x}{\log 12} + \left(\frac{1}{(1-\gamma)\log 2} - \frac{2}{\log 12}\right)(\log x - a\log 12) + \frac{\kappa_L}{(1-\gamma)\log 2}\end{aligned}$$

giving (2.4) from (A.1); similarly, we have

$$\begin{aligned}\frac{2\nu_3(\lceil x\rceil_L^{\langle 2,3\rangle}) - \gamma\nu_2(\lceil x\rceil_L^{\langle 2,3\rangle})}{1-\gamma} &= \frac{2(a+c)-\gamma(2a+b)}{1-\gamma}\\ &\leq 2a + \frac{2(\log x - a\log 12 + \kappa_L)}{(1-\gamma)\log 3}\\ &= \frac{2\log x}{\log 12} + \left(\frac{2}{(1-\gamma)\log 3} - \frac{2}{\log 12}\right)(\log x - a\log 12) + \frac{2\kappa_L}{(1-\gamma)\log 3}\end{aligned}$$

giving (2.5) from (A.1).

To prove parts (i) and (ii) of Lemma 2.1, we establish the following lemma to upper bound κ_L.

Lemma A.1. *If n_1, n_2, m_1, m_2 are natural numbers such that $n_1+n_2, m_1+m_2 \geq 1$ and*

$$1 \leq \frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}}$$

then

$$\kappa_{\min(2^{n_1+n_2}, 3^{m_1+m_2})/6} \leq \log\max\left(\frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}}\right).$$

Proof. If $\min(2^{n_1+n_2}, 3^{m_1+m_2})/6 \leq t \leq 2^{n_2-1}3^{m_1-1}$, then we have

$$t \leq 2^{n_2-1}3^{m_1-1} \leq \max\left(\frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}}\right)t, \tag{A.2}$$

so we are done in this case. Now suppose that $t > 2^{n_2-1}3^{m_1-1}$. If we write $\lceil t\rceil^{\langle 2,3\rangle} = 2^n 3^m$ be the smallest 3-smooth number that is at least t, then we must have $n \geq n_2$ or $m \geq m_1$ (or both).

Thus at least one of $\frac{2^{n_1}}{3^{m_1}}2^n3^m$ and $\frac{3^{m_2}}{3^{n_2}}2^n3^m$ is an integer, and is thus at most t by construction. This gives (A.2), and the claim follows. □

Some efficient choices of parameters for this lemma are given in Table 1. For instance, $\kappa_{4.5} \leq \log\frac{4}{3} = 0.28768\dots$ and $\kappa_{40.5} \leq \log\frac{32}{27} = 0.16989\dots$. In fact, since $\lceil 4.5+\varepsilon\rceil^{\langle 2,3\rangle} = 6$ and $\lceil 40.5+\varepsilon\rceil^{\langle 2,3\rangle} = 48$ for all sufficiently small $\varepsilon > 0$, we see that these bounds are sharp (And similarly for the other entries in Table 1); this establishes part (i).

n_1	m_1	n_2	m_2	$\min(2^{n_1+n_2}, 3^{m_1+m_2})/6$	$\log\max(3^{m_1}/2^{n_1}, 2^{n_2}/3^{m_2})$
1	1	**1**	**0**	$1/2 = 0.5$	$\log 2 = 0.69314\dots$
1	**1**	2	1	$2^2/3 = 1.33\dots$	$\log(3/2) = 0.40546\dots$
3	2	**2**	**1**	$3^2/2 = 4.5$	$\log(2^2/3) = 0.28768\dots$
3	2	**5**	**3**	$3^4/2 = 40.5$	$\log(2^5/3^3) = 0.16989\dots$
3	**2**	8	5	$2^{10}/3 = 341.33\dots$	$\log(3^2/2^3) = 0.11778\dots$
11	**7**	8	5	$2^{18}/3 = 87381.33\dots$	$\log(3^7/2^{11}) = 0.06566\dots$
19	12	**8**	**5**	$3^{17}/2 \approx 6.4\times 10^7$	$\log(2^8/3^5) = 0.05211\dots$
19	12	**27**	**17**	$3^{29}/2 \approx 3.4\times 10^{13}$	$\log(2^{27}/3^{17}) = 0.03856\dots$
19	12	**46**	**29**	$3^{41}/2 \approx 1.8\times 10^{19}$	$\log(2^{46}/3^{29}) = 0.02501\dots$

TABLE 1. Efficient parameter choices for Lemma A.1. The parameters used to attain the minimum or maximum are indicated in **boldface**. Note how the number of rows in each group matches the terms $1, 1, 2, 2, 3, \dots$ in the continued fraction expansion.

Remark A.2. It should be unsurprising that the continued fraction convergents $1/1$, $2/1$, $3/2$, $8/5$, $19/12$, … to

$$\frac{\log 3}{\log 2} = 1.5849\dots = [1; 1, 1, 2, 2, 3, 1, \dots]$$

are often excellent choices for n_1/m_1 or n_2/m_2, although other approximants such as $5/3$ or $11/7$ are also usable.

Finally, we establish (ii). From the classical theory of continued fractions, we can find rational approximants

$$\frac{p_{2j}}{q_{2j}} \leq \frac{\log 3}{\log 2} \leq \frac{p_{2j+1}}{q_{2j+1}} \tag{A.3}$$

to the irrational number $\log 3/\log 2$, where the convergents p_j/q_j obey the recursions

$$p_j = b_j p_{j-1} + p_{j-2}; \quad q_j = b_j q_{j-1} + q_{j-2}$$

with $p_{-1} = 1, q = -1 = 0, p_0 = b_0, q_0 = 1$, and

$$[b_0; b_1, b_2, \dots] = [1; 1, 1, 2, 2, 3, 1 \dots]$$

is the continued fraction expansion of $\frac{\log 3}{\log 2}$. Furthermore, $p_{2j+1}q_{2j} - p_{2j}q_{2j+1} = 1$, and hence

$$\frac{\log 3}{\log 2} - \frac{p_{2j}}{q_{2j}} = \frac{1}{q_{2j}q_{2j+1}}. \tag{A.4}$$

By Baker's theorem, $\frac{\log 3}{\log 2}$ is a Diophantine number, giving a bound of the form

$$q_{2j+1} \ll q_{2j}^{O(1)} \tag{A.5}$$

and a similar argument (using $p_{2j+2}q_{2j+1} - p_{2j+1}q_{2j+2} = -1$) gives

$$q_{2j+2} \ll q_{2j+1}^{O(1)}. \tag{A.6}$$

We can rewrite (A.3) as

$$1 \leq \frac{3^{q_{2j}}}{2^{p_{2j}}}, \frac{2^{p_{2j+1}}}{3^{q_{2j+1}}}$$

and routine Taylor expansion using (A.4) gives the upper bounds

$$\frac{3^{q_{2j}}}{2^{p_{2j}}}, \frac{2^{p_{2j+1}}}{3^{q_{2j+1}}} \leq \exp\left(O\left(\frac{1}{q_{2j}}\right)\right).$$

From Lemma A.1 we obtain

$$\kappa_{\min(2^{p_{2j}+p_{2j+1}}, 3^{q_{2j}+q_{2j+1}})/6} \ll \frac{1}{q_{2j}}.$$

The claim then follows from (A.5), (A.6) (and the obvious fact that κ is monotone non-increasing after optimizing in j.

Remark A.3. It seems reasonable to conjecture that c can be taken to be arbitrarily close to 1, but this is essentially equivalent to the open problem of determining that irrationality measure of $\log 3/\log 2$ is equal to 2.

Appendix B. Estimating sums over primes

In this appendix we establish Lemma 2.2. The key tool is

Lemma B.1 (Integration by parts)**.** *Let $(y, x]$ be a half-open interval in $(0, +\infty)$. Suppose that one has a function $a : \mathbb{N} \to \mathbb{R}$ and a continuous function $f : (y, x] \to \mathbb{R}$ such that*

$$\sum_{y<n\leq z} a_n = \int_z^y f(t)\, dt + C + O_{\leq}(A)$$

for all $y \leq z \leq x$, and some $C \in \mathbb{R}$, $A > 0$. Then, for any function $b : (y, x] \to \mathbb{R}$ of bounded total variation, one has

$$\sum_{y<n\leq x} b(n)a_n = \int_x^y b(t)f(t)\, dt + O_{\leq}(A\|b\|_{\mathrm{TV}^*(y,x]}). \tag{B.1}$$

Proof. If, for every natural number $y < n \leq x$, one modifies b to be equal to the constant $b(n)$ in a small neighborhood of n, then one does not affect the left-hand side of (B.1) or increase the total variation of b, while only modifying the integral in (B.1) by an arbitrarily small amount. Hence, by the usual limiting argument, we may assume without loss of generality that b is locally constant at each such n. If we define the function $g : (y, x] \to \mathbb{R}$ by

$$g(z) := \sum_{y<n\leq z} a_n - \int_z^y f(u)\, du - C$$

then g has jump discontinuities at the natural numbers, but is otherwise continuously differentiable, and is also bounded uniformly in magnitude by A. We can then compute the Riemann–Stieltjes integral

$$\int_{(y,x]} b\, dg = \sum_{y<n\leq x} b(n)a_n - \int_y^x f(t)b(t)\, dt.$$

Since the discontinuities of g and b do not coincide, we may integrate by parts to obtain

$$\int_{(y,x]} b\, dg = b(x)g(x) - b(y^+)g(y^+) - \int_{(y,x]} g db.$$

The left-hand side is $O_{\leq}(A\|b\|_{\mathrm{TV}^*(y,x]})$, and the claim follows. □

We now prove (2.14). In fact we prove the sharper estimate

$$\sum_{y<p\leq x} b(p)\log p = \int_y^x b(t)(1-\frac{2}{\sqrt{t}})\, dt + O_{\leq}\left(\|b\|_{\mathrm{TV}^*((y,x])}\tilde{E}(x)\right) \tag{B.2}$$

where

$$\tilde{E}(x) := 0.95\sqrt{x} + \min(\max(\varepsilon_0, \varepsilon_1(x)), \varepsilon_2(x), \varepsilon_3(x))1_{x\geq 10^{19}} \tag{B.3}$$

and

$$\begin{aligned}
\varepsilon_0(x) &:= \frac{\sqrt{x}}{8\pi}\log x(\log x - 3)\\
\varepsilon_1(x) &:= 1.12494\times 10^{-10}\\
\varepsilon_2(x) &:= 9.39(\log^{1.515} x)\exp(-0.8274\sqrt{\log x})\\
\varepsilon_3(x) &:= 0.026(\log^{1.801} x)\exp(-0.1853(\log^{3/5} x)(\log\log x)^{-1/5})
\end{aligned}$$

From using the ε_2 term, it is clear that

$$\tilde{E}(x) \ll x\exp(-c\sqrt{\log x})$$

for some absolute constant $c > 0$; and by using the $\varepsilon_0, \varepsilon_1$ term and routine calculations one can show that

$$\tilde{E}(x) \leq E(x)$$

for all $x \geq 1423$.

Observe that $\tilde{E}$ is monotone non-decreasing. Thus by Lemma B.1, to show (B.2) will suffice to show that

$$\sum_{p\leq x}\log p = x - \sqrt{x} + O_{\leq}(\tilde{E}(x)) = \int_0^x \left(1-\frac{2}{\sqrt{t}}\right) dt + O_{\leq}(\tilde{E}(x))$$

for all $x \geq 1423$.

For $1423 \leq x \leq 10^{19}$, this claim follows from [4, Theorem 2]. For $x > 10^{19}$, we apply [3, (6.10), (6.11)] to conclude that

$$\sum_{p\leq x}\log p = \psi(x) - \psi(\sqrt{x}) + O_{\leq}(1.03883(x^{1/3} + x^{1/5} + 2(\log x)x^{1/13})),$$

where $\psi(x) := \sum_{n\leq x} \Lambda(n)$ is the usual von Mangoldt summatory function. From [14, Theorems 10,12] we have

$$\psi(\sqrt{x}) = \sqrt{x} + O_{\leq}(0.18\sqrt{x}).$$

Since

$$0.18\sqrt{x} + 1.03883(x^{1/3} + x^{1/5} + 2(\log x)x^{1/13}) \leq 0.95\sqrt{x}$$

in this range of x, it suffices to show that

$$\psi(x) = x + O_{\leq}(\min(\max(\varepsilon_0(x), \varepsilon_1(x)), \varepsilon_2(x), \varepsilon_3(x)))$$

for $x > 10^{19}$. The claims for $i = 2, 3$ follow from [11, Theorems 1.1, 1.4]. In [3, Theorem 2, (7.3)], the bound

$$\psi(x) = x + O_{\leq}(\varepsilon_0(x))$$

is established whenever $x \geq 5000$ and $4.92\frac{x}{\sqrt{\log x}} \leq T$, where T is a height up to which the Riemann hypothesis has been established. Using the value $T = 3 \times 10^{12}$ from [12], we can therefore cover the range $10^{19} < x < e^{55}$ (in fact we could go up to $e^{58.33} \approx 2.1 \times 10^{25}$). For $x \geq e^{55}$, we can use [3, Table 2] (the value $T = 2.445 \times 10^{12}$ used there following from [12]).

Remark B.2. Assuming the Riemann hypothesis, the $\varepsilon_1, \varepsilon_2, \varepsilon_3$ terms in the definition of $\tilde{E}(x)$ may be deleted, since [3, (7.3)] then holds for all $x \geq 5000$.

The claim (2.16) now follows from (2.14) by setting $b(t) := \frac{1}{\log t}$. For non-negative b, the claims (2.17), (2.18) follow from (2.14) and the pointwise bounds

$$\frac{1}{\log x} b(p) \log p \leq b(p) \leq \frac{1}{\log y} b(p) \log p$$

and

$$1 - \frac{2}{\sqrt{y}} \leq 1 - \frac{2}{\sqrt{t}} \leq 1.$$

Finally, (2.19), (2.20) come from specializing (2.17), (2.18) to the case of an indicator function $b = 1_{(y,x]}$.

Appendix C. Computation of c_0

In this appendix we give some details regarding the rigorous numerical estimation of the constant c_0 defined in (1.6). As one might imagine from an inspection of Figure 3, direct application of numerical quadrature converges quite slowly due to the oscillatory singularity. To resolve the singularity, we can perform a change of variables $x = 1/y$ to express c_0 as an improper integral:

$$c_0 = \frac{1}{e} \int_1^\infty \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2}. \tag{C.1}$$

The integrand is piecewise smooth and the integral can be computed explicitly on any interval $[a, b]$ of the form

$$[a, b] \subset [n, n+1] \cap [(m-1)e, me]$$

for some non-negative integers n, m as

$$\int_a^b \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2} = n \left(\frac{\log(b/m)}{b} - \frac{\log(a/m)}{a} \right).$$

This formula permits one to evaluate $\int_1^b \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2}$ exactly for any finite b. To control the tail, we see from the crude bounds $0 \leq \lfloor y \rfloor \leq y$ and

$$0 \leq \log \frac{\lceil y/e \rceil}{y/e} \leq \log\left(1 + \frac{e}{y}\right) \leq \frac{e}{y}$$

that

$$0 \leq \int_b^\infty \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2} \leq \frac{e}{b} \tag{C.2}$$

which allows for rigorous upper and lower bounds on the improper integral. For instance, this procedure gives

$$0.304419004 \leq c_0 \leq 0.304419017.$$

Heuristically, the tail integral (C.2) should be approximately $e/2b$ due to the equidistribution properties of the fractional part of y/e. Using this heuristic approximation, one obtains the prediction

$$c_0 \approx 0.30441901087.$$

It should be possible to obtain this level of precision more rigorously (using interval arithmetic to preclude any possibility of roundoff error), but we have not attempted to do so.

References

[1] K. Alladi, C. Grinstead, *On the decomposition of $n!$ into prime powers*, J. Number Theory **9** (1977) 452–458.

[2] S. F. Assmann, D. S. Johnson, D. J. Kleitman, J. Y.-T. Leung, *On a dual version of the one-dimensional bin packing problem*, J. Algorithms **5** (1984) 502–525.

[3] J. Büthe, *Estimating $\pi(x)$ and related functions under partial RH assumptions*, Math. Comp., 85(301), 2483–2498, Jan. 2016.

[4] J. Büthe, *An analytic method for bounding $\psi(x)$* . Math. Comp., **87** (312), 1991–2009.

[5] P. Dusart, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018) 227–251.

[6] P. Erdős, *Some problems in number theory*, in Computers in Number Theory, Academic Press, London New York, 1971, pp. 405–414.

[7] P. Erdős, *Some problems I presented or planned to present in my short talk*, Analytic number theory, Vol. 1 (Allerton Park, IL, 1995) (1996), 333–335.

[8] P. Erdős, R. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathematique 1980.

[9] R. K. Guy, *Unsolved Problems in Number Theory*, 3rd Edition, Springer, 2004.

[10] R. K. Guy, J. L. Selfridge, *Factoring factorial n*, Amer. Math. Monthly **105** (1998) 766–767.

[11] D. Johnston, A. Yang, *Some explicit estimates for the error term in the prime number theorem*, J. Math. Anal. Appl., **527** (2) (2023), Paper No. 127460.

[12] D. Platt, T. Trudgian, *The Riemann hypothesis is true up to $3 \cdot 10^{12}$*, Bull. Lond. Math. Soc. **53** (2021), no. 3, 792–797.

[13] H. Robbins, *A Remark on Stirling's Formula*, Amer. Math. Monthly **62** (1955) 26–29.

[14] J. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6 (1962), 64–94.

[15] T. Tao, *Decomposing factorials into bounded factors*, preprint, 2025. `https://arxiv.org/abs/2503.20170v2`

[16] T. Tao, *Verifying the Guy–Selfridge conjecture*, Github repository, 2025. `https://github.com/teorth/erdos-guy-selfridge`.

???

Email address: ???

???

Email address: ???

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1555.

Email address: tao@math.ucla.edu

???

Email address: ???