

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

### On the number of arithmetic formulas

**This is a pre print version of the following article:**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1622119> since 2017-01-16T17:33:33Z

*Published version:*

DOI:10.1142/S1793042115500591

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# ON THE NUMBER OF ARITHMETIC FORMULAS

CARLO SANNA

ABSTRACT. For each positive integers  $n$ , let  $g(n)$  be the number of arithmetic expressions evaluating to  $n$  and involving only the constant 1, additions and multiplications, with the restriction that multiplication by 1 is not allowed. We consider two arithmetic expressions to be equal if one can be obtained from the other through a repeated application of the commutative and associative properties. We give an algorithm to compute  $g(n)$  and prove that  $\log(g(n)) = \beta n + O(\sqrt{n})$ , as  $n \rightarrow +\infty$ , where  $\beta := \log(24)/24$ .

## 1. INTRODUCTION

Given a positive integer  $n$ , a quite natural question is to study the arithmetic expressions only using the constant 1 and the binary operators of addition and multiplication, which return  $n$  as result. Since 1 is the multiplicative identity, we restrict ourselves to consider only expressions where 1 is never an argument of multiplication. We call these expressions *arithmetic formulas* for  $n$ . For example, 4 has 6 arithmetic formulas, namely

$$\begin{aligned} &1 + (1 + (1 + 1)), \quad 1 + ((1 + 1) + 1), \quad (1 + (1 + 1)) + 1, \\ &((1 + 1) + 1) + 1, \quad (1 + 1) + (1 + 1), \quad (1 + 1) \times (1 + 1). \end{aligned}$$

Let  $f(n)$  be the number of arithmetic formulas for  $n$ . Gnang, Radziwiłł and the author [5] proved the following asymptotic formula for  $f(n)$ , previously conjectured by Gnang and Zeilberger [6].

**Theorem 1.1.** *There exists constants  $C > 0$  and  $\rho > 4$  such that*

$$f(n) \sim C \cdot \frac{\rho^n}{n^{3/2}},$$

as  $n \rightarrow \infty$ .

Clearly, given an arithmetic formula for  $n$ , if we apply to it, or to one of its subformulas, the commutative or associative identities:

$$\begin{aligned} &A + B = B + A, \quad A \times B = B \times A, \\ &(A + B) + C = A + (B + C), \quad (A \times B) \times C = A \times (B \times C); \end{aligned}$$

then we get another arithmetic formula for  $n$ . We call such a pair of arithmetic formulas *equivalent*, since obviously this is an equivalence relation. To continue the example, 4 has only 2 inequivalent arithmetic formulas, namely

$$1 + 1 + 1 + 1, \quad (1 + 1) \times (1 + 1).$$

Let  $g(n)$  be the number of inequivalent arithmetic formulas for  $n$ .

It is quite reasonable to expect that  $g(n)$  is much smaller than  $f(n)$  as  $n \rightarrow +\infty$ . In fact, we prove the following

**Theorem 1.2.** *It results*

$$\log g(n) = \frac{\log 24}{24} \cdot n + O(\sqrt{n}),$$

as  $n \rightarrow +\infty$ .

---

2010 *Mathematics Subject Classification.* Primary: 05A16, 11A67, 11B75. Secondary: 11A99, 11Y16.  
*Key words and phrases.* Arithmetic formulas, arithmetic expressions, partitions, asymptotic formulas.

## 2. PRELIMINARIES

In this section, we collect some lemmas needed in the proof of Theorem 1.2. For each positive integer  $n$ , let  $p(n)$  be the partition function of  $n$ , i.e., the number of ways of writing  $n$  as a sum of positive integers, where the order of the summands does not matter. Hardy and Ramanujan [7], and independently Uspensky [12], proved the asymptotic formula

$$p(n) \sim \frac{e^{c_0\sqrt{n}}}{4\sqrt{3}n},$$

as  $n \rightarrow +\infty$ , where  $c_0 := \pi\sqrt{2/3}$ . This result was improved by Rademacher [10, 11] who gave a convergent series expression for  $p(n)$ . We only need the following upper bound.

**Lemma 2.1.** *For each positive integer  $n$ , we have  $p(n) < e^{c_0\sqrt{n}}$ .*

*Proof.* See [1][Theorem 14.5]. □

The multiplicative analogue of the (additive) partition function  $p(n)$  is the multiplicative partition function  $q(n)$ , i.e., the number of ways of writing  $n$  as a product of integers larger than 1, where the order of the factors is not taken into account. The maximal order of  $q(n)$  was studied by Canfield, Erdős and Pomerance [2] (correcting [9]) and at that time an effective upper bound for  $q(n)$  was given by Hughes and Shallit [8]. We make use of the following

**Lemma 2.2.** *For each positive integer  $n$ , we have  $q(n) \leq n$ .*

*Proof.* See [4]. □

The last two lemmas of this section are more technical. The first is a recursive formula for a sum over the additive partitions of  $n$ . The second concerns two inequalities for multiplicative partitions of  $n$ .

**Lemma 2.3.** *Let  $f$  be a function from the positive integers to the complex numbers and define*

$$F(n) := \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k}} f(a_1) \cdots f(a_k),$$

*for all positive integers  $n$ , put also  $F(0) := 1$ . Then*

$$F(n) = \frac{1}{n} \sum_{m=1}^n F(n-m) h(m),$$

*for all positive integers  $n$ , where*

$$h(m) := \sum_{d|m} (f(d))^{m/d} d.$$

*Proof.* This result is a generalization of the classical identity for the partition function

$$(1) \quad p(n) := \frac{1}{n} \sum_{m=1}^n p(n-m) \sigma(m),$$

where  $\sigma(m) := \sum_{d|m} d$  is the sum of divisors of  $m$ . A detailed proof of (1) that can be easily adapted to this generalization can be found in [3, Theorem 3]. □

**Lemma 2.4.** *Let  $a_1, \dots, a_k \geq 2$  be integers, with  $k \geq 2$ , and set  $n := a_1 \cdots a_k$ . Then*

$$(2) \quad a_1 + \cdots + a_k \leq 2 + \frac{n}{2}.$$

*Moreover, if  $n \geq 129$  then*

$$(3) \quad \sqrt{a_1} + \cdots + \sqrt{a_k} \leq \sqrt{2} + \sqrt{\frac{n}{2}}.$$

*Proof.* We proceed by induction on  $k$ . If  $k = 2$  consider that for each  $t > 0$  the map  $f_t : x \mapsto x^t + (n/x)^t$  is derivable on  $[2, n/2]$ . Moreover,  $f'_t(x) < 0$  for  $x \in [2, \sqrt{n}[$ ,  $f'_t(x) > 0$  for  $x \in ]\sqrt{n}, n/2]$  and  $f_t(2) = f_t(n/2) = 2^t + (n/2)^t$ . Therefore, it results  $f'_t(x) \leq 2^t + (n/2)^t$ , for each  $x \in [2, n/2]$ . Thus, taking  $t = 1$  and  $t = 1/2$  we get

$$a_1 + a_2 \leq 2 + \frac{n}{2} \quad \text{and} \quad \sqrt{a_1} + \sqrt{a_2} \leq \sqrt{2} + \sqrt{\frac{n}{2}},$$

respectively. Suppose  $k \geq 3$  and that the claim holds for all integers in  $[2, k-1]$ . Since  $a_1, \dots, a_k \geq 2$ , we have

$$a_1 + \dots + a_k \leq a_1 \cdots a_{k-1} + a_k \leq 2 + \frac{n}{2},$$

by inductive hypothesis, so (2) is proved. The inequality (3) needs a bit more work. If  $k \geq 8$  then

$$\frac{\sqrt{a_1} + \dots + \sqrt{a_k}}{\sqrt{n}} = \sum_{i=1}^k \frac{1}{\sqrt{a_1 \cdots a_i \cdots a_k}} \leq \frac{k}{\sqrt{2^{k-1}}} \leq \frac{1}{\sqrt{2}},$$

and the claim follows. Thus we can assume  $k \leq 7$  and, without loss of generality,  $a_1 \leq \dots \leq a_k$ . If  $a_{k-1} \geq 4$  or  $a_k \geq 12$  then

$$\sqrt{a_1} + \dots + \sqrt{a_{k-1}} + \sqrt{a_k} \leq \sqrt{a_1} + \dots + \sqrt{a_{k-2}} + \sqrt{a_{k-1}a_k} \leq \sqrt{2} + \sqrt{\frac{n}{2}},$$

by inductive hypothesis. Hence suppose  $a_{k-1} \leq 3$  and  $a_k \leq 11$ , so that  $129 \leq n \leq 3^{k-1} \cdot 11$  and then  $k \geq 4$ . Now if  $a_{k-2} \geq 3$  then

$$\sqrt{a_1} + \dots + \sqrt{a_{k-2}} + \sqrt{a_{k-1}} + \sqrt{a_k} \leq \sqrt{a_1} + \dots + \sqrt{a_{k-3}} + \sqrt{a_{k-2}a_{k-1}a_k} \leq \sqrt{2} + \sqrt{\frac{n}{2}},$$

by inductive hypothesis. So we can assume  $4 \leq k \leq 7$ ,  $a_i = 2$  for  $i \leq k-2$ ,  $a_{k-1} \leq 3$  and  $a_k \leq 11$ . At this point, it is easy to check that (3) holds for all possible values of  $a_1, \dots, a_k$  satisfying  $n \geq 129$ .  $\square$

Note that for any even integer  $n \geq 4$ , taking  $k = 2$ ,  $a_1 = 2$  and  $a_2 = n/2$  we have the equality in both (2) and (3). Furthermore, if  $n = 128$ ,  $k = 7$  and  $a_1 = \dots = a_7 = 2$  then (3) does not hold.

### 3. INEQUIVALENT ARITHMETIC FORMULAS

In [5] arithmetic formulas are defined as special plane binary trees. To study inequivalent arithmetic formulas is better instead to define them as trees, in the following way

**Definition 3.1.** Let  $n$  be a positive integer. An inequivalent arithmetic formula for  $n$  is an  $\mathbf{N}$ -valued  $\{+, \times\}$ -labelled rooted tree such that

- (i) The value of the root is  $n$ .
- (ii) The value of each leaf is 1.
- (iii) Each non-leaf node has at least two children.
- (iv) Each non-leaf node is labelled with  $+$  (additive node) or  $\times$  (multiplicative node).
- (v) Each child of an additive node is a leaf or a multiplicative node.
- (vi) Each child of a multiplicative node is an additive node.
- (vii) The value of each additive node is the sum of the values of its children.
- (viii) The value of each multiplicative node is the product of the values of its children.

Now it is not difficult to check that this definition is consistent with the one given in the Introduction. Precisely, there is a bijection between the set of inequivalent arithmetic formulas and the set of equivalence classes of arithmetic formulas, so  $g(n)$  is actually the number of inequivalent arithmetic formulas, with respect to Definition 3.1.

To study  $g(n)$  we need to introduce two auxiliary functions: For each integer  $n \geq 2$  let  $g^+(n)$ , respectively  $g^\times(n)$ , be the number of arithmetic formulas for  $n$  which root node is additive,

respectively multiplicative; set also  $g^+(1) := 0$  and  $g^\times(1) := 1$ . Thus it is straightforward that  $g(n) = g^+(n) + g^\times(n)$  for all positive integers  $n$ . Furthermore, we have the following recursive formulas for computing  $g^+(n)$  and  $g^\times(n)$ .

**Lemma 3.1.** *For all integers  $n \geq 2$ ,*

$$g^+(n) = \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k < n}} g^\times(a_1) \cdots g^\times(a_k)$$

and

$$g^\times(n) = \sum_{\substack{n=a_1 \cdots a_k \\ 2 \leq a_1 \leq \dots \leq a_k \leq n/2}} g^+(a_1) \cdots g^+(a_k).$$

For “small”  $n$ , we can actually use the formulas of Lemma 3.1 to find  $g(n)$ . But since the formula for  $g^+(n)$  involve a sum over all the  $p(n)$  partitions of  $n$ , this is not a practical method for “large”  $n$ . Our proof of Theorem 1.2 requires the computation of  $g(n)$  for all  $n \leq 1052$ , so we need a better algorithm.

**Lemma 3.2.** *For all positive integers  $n$ , we have*

$$g(n) = \frac{1}{n} \sum_{m=1}^n g(n-m) h(m),$$

where  $g(0) := 1$  and

$$h(m) := \sum_{d|m} (g^\times(d))^{m/d} d.$$

*Proof.* Note that

$$\begin{aligned} g(n) &= g^\times(n) + g^+(n) = g^\times(n) + \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k < n}} g^\times(a_1) \cdots g^\times(a_k) \\ &= \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k}} g^\times(a_1) \cdots g^\times(a_k), \end{aligned}$$

so the claim follows from Lemma 2.3 with  $f = g^\times$ .  $\square$

Putting together Lemma 3.1 and Lemma 3.2, we get the following pseudocode of an algorithm to compute  $g(n)$ ,  $g^+(n)$  and  $g^\times(n)$  for  $n \leq N$ .

$$g(0) = 1, g(1) = 1, g^+(1) = 0, g^\times(1) = 1, h(1) = 1$$

**for**  $n = 2, \dots, N$  **do**

$$g^\times(n) = \sum_{n=a_1 \cdots a_k, 2 \leq a_1 \leq \dots \leq a_k} g^+(a_1) \cdots g^+(a_k)$$

$$h(n) = \sum_{d|n} (g^\times(d))^{n/d} d$$

$$g(n) = (1/n) \sum_{m=1}^n g(n-m) h(m)$$

$$g^+(n) = g(n) - g^\times(n)$$

**end for**

The author implemented this algorithm in the Python programming language, since it has native support for arbitrary sized integers. Running the implementation on a personal computer, in few seconds he found  $g(n)$ ,  $g^+(n)$  and  $g^\times(n)$  for each  $n \leq N$ , with  $N$  of the order of  $10^3$ .

#### 4. PROOF OF THEOREM 1.2

The proof of Theorem 1.2 consists of two lemmas. From now on, let  $\beta := \frac{\log 24}{24}$ .

**Lemma 4.1.** *For all positive integers  $n$ , we have*

- (i).  $g^\times(n) \leq \exp(\beta n)$ , with equality for  $n = 24$ ; and  
(ii).  $g(n) < \exp(\beta n + c_0\sqrt{n})$ .

*Proof.* We prove (i) by strong induction on  $n$ . With the aid of a computer, using the algorithm of the previous section, one can check that  $g^\times(n) \leq \exp(\beta n)$  holds for each positive integer  $n \leq 1052$ . Therefore, let  $m \geq 1053$  be an integer and suppose that  $g^\times(n) \leq \exp(\beta n)$  for all positive integers  $n < m$ . By inductive hypothesis, from Lemma 3.1 and Lemma 2.1 it follows that

$$(4) \quad \begin{aligned} g(n) &= \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k}} g^\times(a_1) \cdots g^\times(a_k) \leq \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k}} e^{\beta(a_1+\dots+a_k)} \\ &= p(n) e^{\beta n} < e^{\beta n + c_0\sqrt{n}}, \end{aligned}$$

for all positive integers  $n < m$ . As a consequence, using Lemmas 3.1, 2.4 and 2.2 we obtain

$$\begin{aligned} g^\times(n) &= \sum_{\substack{n=a_1 \cdots a_k \\ 2 \leq a_1 \leq \dots \leq a_k \leq n/2}} g^+(a_1) \cdots g^+(a_k) \leq \sum_{\substack{n=a_1 \cdots a_k \\ 2 \leq a_1 \leq \dots \leq a_k \leq n/2}} g(a_1) \cdots g(a_k) \\ &< \sum_{\substack{n=a_1 \cdots a_k \\ 2 \leq a_1 \leq \dots \leq a_k \leq n/2}} e^{\beta(a_1+\dots+a_k) + c_0(\sqrt{a_1}+\dots+\sqrt{a_k})} \\ &\leq q(n) \exp\left(\frac{1}{2}\beta n + 2\beta + c_0\sqrt{2} + c_0\sqrt{n/2}\right) \\ &\leq \exp\left(\frac{1}{2}\beta n + 2\beta + c_0\sqrt{2} + c_0\sqrt{n/2} + \log n\right), \end{aligned}$$

for all integers  $n \in [129, 2m[$ . In particular, for  $n = m$  we have

$$g^\times(m) \leq \exp\left(\frac{1}{2}\beta m + 2\beta + c_0\sqrt{2} + c_0\sqrt{m/2} + \log m\right) \leq \exp(\beta m),$$

where the last inequality follows since  $m \geq 1053$ , hence (i) is proved. Now (ii) follows from (4) and the proof is complete.  $\square$

**Lemma 4.2.** *We have  $g^+(n) \geq \exp(\beta n)$  for all integers  $n \geq 7$ .*

*Proof.* The claim is true for  $n \in [7, 30]$ , so assume  $n \geq 31$ . We can write  $n = 24q + r$ , for some integers  $q \geq 1$  and  $r \in [7, 30]$ . From Lemma 4.1 we know that  $g^\times(24) = e^{24\beta}$ , hence using Lemma 3.1 we get

$$\begin{aligned} g^+(n) &= \sum_{\substack{n=a_1+\dots+a_k \\ a_1 \leq \dots \leq a_k < n}} g^\times(a_1) \cdots g^\times(a_k) \geq (g^\times(24))^q \sum_{\substack{r=b_1+\dots+b_h \\ b_1 \leq \dots \leq b_h < r}} g^\times(b_1) \cdots g^\times(b_h) \\ &= (g^\times(24))^q \cdot g^+(r) \geq \exp(\beta(24q + r)) = \exp(\beta n), \end{aligned}$$

which is our claim.  $\square$

At this point, Theorem 1.2 is an easy consequence of Lemma 4.2 and point (ii) of Lemma 4.1, since together they imply

$$\exp(\beta n) \leq g^+(n) \leq g(n) < \exp(\beta n + c_0\sqrt{n}),$$

for all integers  $n \geq 7$ . Therefore,  $\log g(n) = \beta n + O(\sqrt{n})$ , as  $n \rightarrow +\infty$ .

## REFERENCES

- [1] T. M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York, 1976.
- [2] E. R. Canfield, P. Erdős, and C. Pomerance, *On a problem of Oppenheim concerning "factorisatio numerorum"*, J. Number Theory **17** (1983), 1–28.
- [3] T. R. Chandrupatla, A. Hassen, and T. Osler, *Surprising connections between partitions and divisions*, College Math. J. **38** (2007), no. 4, 278–287.
- [4] F. W. Dodd and L. E. Mattics, *A bound for the number of multiplicative partitions*, Amer. Math. Monthly **93** (1986), 125–126.
- [5] E. K. Gngang, M. Radziwill, and C. Sanna, *Counting arithmetic formulas*, European J. Combin. **47** (2015), 40–53.

- [6] E. K. Gnang and D. Zeilberger, *Zeroless arithmetic: Representing integers ONLY using ONE*, J. Difference Equ. Appl. **19** (2013), no. 11, 1921–1926.
- [7] G. H. Hardy and S. Ramanujan, *Asymptotic formulae in combinatory analysis*, Proc. London Math. Soc. **17** (1918), 75–115.
- [8] J. F. Hughes and J. O. Shallit, *On the number of multiplicative partitions*, Amer. Math. Monthly **90** (1983), 468–471.
- [9] A. Oppenheim, *On an arithmetic function*, J. London Math. Soc. **1** (1926), 205–211.
- [10] H. Rademacher, *On the partition function  $p(n)$* , Proc. London Math. Soc. **43** (1937), 241–254.
- [11] H. Rademacher, *On the expansion of the partition function in a series*, Ann. Math. **44** (1943), 416–422.
- [12] Ya. V. Uspensky, *Asymptotic expressions of numerical functions occurring in problems concerning the partition of numbers into summands*, Bull. Acad. Sci. de Russie **14** (1920), 199–218.

UNIVERSITÀ DEGLI STUDI DI TORINO, DEPARTMENT OF MATHEMATICS, TURIN, ITALY  
E-mail address: `carlo.sanna.dev@gmail.com`