Multilayer Factorization of Catalan Numbers

Gennady Eremin
argenns@gmail.com

July 12, 2016

Abstract. The article describes a Prime Factorization of Catalan numbers. Odd prime factors are distributed in layers. In each layer the noncrossing segments contain only non-repeated prime numbers. Repeated factors are formed when primes are duplicated among different layers.

Key Words: Prime factorization, decomposition, Catalan numbers, modular arithmetic, segment, Legendre.

1. Introduction

It is currently very hard to factorize large integers. Much more difficult to factorize a giant number if it is not given in natural form. This is what happens with special numbers, i.e. items of known numerical sequences. For example, the *million*th Catalan number is 600,000-digit integer, and we don't know its natural form. Special numbers are characterized by mutual relationships between elements of the corresponding sequences, and it simplifies a Prime Factorization.

The Catalan numbers appear in various combinatorial applications (see [Stan15]). The explicit formula of the general term of the Catalan sequence is

$$C(n) = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}, \quad n \geq 0. \tag{1.1}$$

The first Catalan numbers are 1, 1, 2, 5, 14, 42, 132, 429, 1430, … (see [A000108]). We are interested in composite Catalan numbers, i.e. $C(n) > 5$. Therefore, in the future $n > 3$. The formula (1.1) can be converted to a more convenient equation for practical calculations (see [Wei16])

$$C(n) = 2^n \times (2n-1)!!\big/(n+1)!, \tag{1.1a}$$

where $(2n-1)!! = 1\cdot3\cdot5\cdots(2n-1)$ is a double odd factorial.

Let $\mathbb{R}$ denote the real numbers. From (1.1a) immediately follows:

(a) Prime factors of $C(n)$ are less than $2n$, i.e. these primes are contained in the open interval $B = (1, 2n) \subset \mathbb{R}$. Let us call B the *factor-base* of $C(n)$.

(b) Any prime $p \in (n+1, 2n) = S_1$ divides $C(n)$ or $p \mid C(n)$. Let us call S_1 the *segment* of B (usually there are other smaller segments).

(c) Any prime $p \in S_1$ is distinct (non-repeated) factor of $C(n)$, i.e. $p \mid C(n)$, but $p^2 \nmid C(n)$.

The segment S_1 is usually not the only one in factor-base B. In the next segment $S_2 = (\frac{1}{2}(n+1), \frac{2}{3}n)$ all primes also divide $C(n)$, and these primes are distinct too. Between S_1 and S_2, there is the close *no-go zone* $[\frac{2}{3}n, n+1]$. It is easy to show that all prime in this no-go zone do not divide $C(n)$.

The article [Er16] deals with segments within the open interval $B_o = (\sqrt{2n}, 2n)$, which covers the overwhelming part of the factor-base B. In B_o there are only distinct prime factors. While the initial tiny interval $(1, \sqrt{2n})$ includes prime factor and all powers of prime factors of $C(n)$. The general form of the kth segment within B_o is the following:

$$S_k = \left(\frac{n+1}{k},\ \frac{2n}{2k-1}\right),\quad k \geq 1. \tag{1.2}$$

Proposition 1.1. *Let* $p > \sqrt{2n}$ *be a prime number and let* $p \mid C(n)$. *Then* $p \in S_i$, $i = \lceil (n+1)/p \rceil$.

This paper proposes the fragmentation of prime factors of a Catalan number on *layers* L_j. In this case each layer contains only unique (non-recurring) primes. Layering often does not correspond to powers of prime factors. For example, for the *millionth* Catalan number the distinct prime factor 101 allocated to the third level, not on the first as expected.

We can say that the collection of layers is a pyramid. The bottom layer L_1 is the most numerous; it selects primes by the segments (1.2). So any prime $p > \sqrt{2n}$ belongs to L_1, if $p \mid C(n)$. Additionally in L_1 there may be some prime factors outside B_o (for example, separate instances of repeated factors). Let us call L_1 a *distinct-layer*, as it accumulates the absolute majority of distinct prime factors of the Catalan number.

Many prime factors (distinct and recurring) are not included in distinct-layer. For any Catalan number, we will prove that $3 \notin L_1$. Let us $B \setminus B_o \setminus \{\sqrt{2n}\}$ call a *main core* of $C(n)$ (see [Er16]).

In this paper we also consider in detail the next *square-layer* L_2. The square-layer accumulates the majority of prime factors that have the power 2. The primes of L_2 are distributed in the truncated factor-base $B^{(2)} = (1, \sqrt{2n})$, the main core, and mostly of these primes fall in the interval $B_o^{(2)} = (\sqrt[3]{2n}, \sqrt{2n})$. In the factor-base $B^{(2)}$ we can select the *second core* $B^{(2)} \setminus B_o^{(2)} \setminus \{\sqrt[3]{2n}\} = (1, \sqrt[3]{2n}) = B^{(3)}$. The general form of the kth segment in $B^{(2)}$ is the following:

$$S_k^{(2)} = \left(\sqrt{\frac{n+1}{k}}, \sqrt{\frac{2n}{2k-1}}\right),\quad k \geq 1.$$

Proposition 1.2. *Let* $p > \sqrt[3]{2n}$ *be a prime number and let* $p^2 \mid C(n)$. *Then* $p \in S_i^{(2)}$, $i = \lceil (n+1)/p^2 \rceil$.

For the nth Catalan number, any distinct prime factor less than $\sqrt[3]{2n}$ can belong to either the distinct-layer or the square-layer or none of them. Usually, it is sufficient to use segments at several (two or three) lower layers. You can calculate primes

only in a tiny core of some truncated factor-base. In addition, segments with large indices are often "empty", i.e. rarely contain prime numbers.

In the general case for *j*th layer we have

$$S_k^{(j)} = \left(\sqrt[j]{\frac{n+1}{k}},\ \sqrt[j]{\frac{2n}{2k-1}} \right),\ k \geq 1,\ j \geq 1. \qquad (1.3)$$

Proposition 1.3. *Let* $p > \sqrt[j+1]{2n}$ *be a prime number and let* $p^j \mid C(n)$. *Then* $p \in S_i^{(j)}$, $i = \lceil (n+1)/p^j \rceil$.

Let us consider some functions which will clarify the situation with a decomposition of prime factors of a Catalan number on layers.

Prime Power Function is often found in the literature. For a prime p and a positive integer m, let $v_p(m)$ denotes the largest power of p dividing m (see [Pom13]). For example, $v_7(14)=1$, $v_7(98)=2$, $v_7(6!)=0$. This function is extended to positive rational numbers as follows:

$$v_p(a/b) = v_p(a) - v_p(b),\ \text{if}\ v_p(a) \geq v_p(b).$$

Here the well-known Legendre's formula

$$v_p(m!) = \sum_{i>0} \lfloor m/p^j \rfloor .$$

According to the equation (1.1), we have for each prime p

$$v_p(C(n)) = \sum_{j>0} \left(\lfloor 2n/p^j \rfloor - \lfloor n/p^j \rfloor - \lfloor (n+1)/p^j \rfloor \right).$$

The above equality explains the partitioning of the factor set on layers. The *j*th non-zero term in the sum is distributed at the *j*th layer of the partitioning. Let's call this procedure *Legendre's decomposition*.

We can obtain a more simplified expression. The power of the odd prime p in the equation (1.1a) is

$$v_p(C(n)) = v_p((2n-1)!!) - \sum_{j>0} \lfloor (n+1)/p^j \rfloor \geq 0,\quad p > 2. \qquad (1.4)$$

Odd Floor Function (round down to the odd number) is useful to calculate the double odd factorial (see [Er16]). For a real $x \geq 0$, let $\lfloor x \rfloor_{\text{odd}}$ denote the floor function to the nearest odd integer, i.e. the fractional part of x is discarded with a decrease to 1 if the result is even or zero. Let's call this operation an *odd floor function*. For example, $\lfloor 23/7 \rfloor_{\text{odd}} = 3$, $\lfloor 31/7 \rfloor_{\text{odd}} = 3$, and $\lfloor 5/7 \rfloor_{\text{odd}} = -1$.

Let $p > 2$ be a prime, let $m > p$ be an odd integer, and let $k = \lfloor m/p \rfloor_{\text{odd}}$. Then

$$v_p(m!!) = v_p((kp)!! \times (kp+2) \times (kp+4) \times \ldots \times m) = v_p((kp)!!) = \tfrac{1}{2}(k+1) + v_p(k!!).$$

By iterating (see [Ep15]), we get the equation for the first operand in (1.4)

$$v_p((2n-1)!!) = \tfrac{1}{2}\sum_{j>0}(\lfloor(2n-1)/p^j\rfloor_{\text{odd}} + 1). \qquad (1.5)$$

Note. This article appeared due to the well-known Kummer's theorem, which deals with the prime factorization of binomial coefficients (see [Pom15]). We will use the methods of Kummer's theorem for the factorization of Catalan numbers.

2. Distinct-layer

Let us give a practical definition for the distinct-layer, which contains the bulk of non-repeated prime factors of the Catalan number. In (1.4-1.5) only the first terms in the sums give rise to distinct-layer items. Let us retain the first summands, i.e. $j=1$. Then for a prime $p>2$ we get the *distinct-layer relation*

$$\tfrac{1}{2}\left(\lfloor(2n-1)/p\rfloor_{\text{odd}}+1\right) - \lfloor(n+1)/p\rfloor = \begin{cases} 1 & \text{if } p \in L_1, \\ 0 & \text{if } p \notin L_1. \end{cases} \qquad (2.1)$$

A binary result in (2.1) is logical, since an arbitrary prime number either belongs to the distinct-layer or not. There are no other options. It is easy to check that (2.1) is equal to 1 for $p \in (n+1, 2n)$.

Let's examine (2.1). For a real number x, let $\{x\} = x - \lfloor x\rfloor$, the fractional part of x. (Next we will try to distinguish a fractional part $\{x\}$ and a singleton set $\{x\}$.) So

$$\lfloor(n+1)/p\rfloor = (n+1)/p - \{(n+1)/p\}.$$

It's more complicated with the odd floor function. The value of $\lfloor(2n-1)/p\rfloor$ may be ***an even number*** (even floor) or ***an odd number*** (odd floor). Zero is impossible because $p<2n$. Let's consider both cases.

Case 1.

$$\lfloor(2n-1)/p\rfloor \text{ is an even number.} \qquad (2.2)$$

Let us convert (2.1):

$$\begin{aligned} &\tfrac{1}{2}\left(\lfloor(2n-1)/p\rfloor - 1 + 1\right) - \left((n+1)/p - \{(n+1)/p\}\right) \\ &\quad = \tfrac{1}{2}\left((2n-1)/p - \{(2n-1)/p\}\right) - n/p - 1/p + \{(n+1)/p\} \\ &\quad = \{(n+1)/p\} - 1/p - \tfrac{1}{2}\left(\{(2n-1)/p\} + 1/p\right) = 0\ (!). \end{aligned} \qquad (2.3)$$

After the cuts we have the natural equality. Zero is logical, since we got the value guaranteed less than 1, and a negative result is impossible. Still, it is desirable to prove a null result for the "clean conscience".

Let us analyze the equation (2.3) using modular arithmetic (e.g., see [Adam05]). First, in (2.3) the factor $\frac{1}{2}$ is valid if and only if

$$(2n-1) \bmod p \;=\; r \text{ is an odd integer,}$$

i.e. $(2n-1)$ is being divided by p gives an odd remainder $r \le p-2$. And this corresponds to the equation

$$a = p \times q + r.$$

In other words, the odd dividend $a = 2n-1$ is being divided by the odd divisor p leaves an odd (not zero!) remainder r if the quotient $q = \lfloor (2n-1)/p \rfloor$ is even. Thus we can write the corresponding operand in (2.3) as follows

$$½\,(\{(2n-1)/p\} + 1/p) = ½\,(\{2n/p\} - 1/p + 1/p) = ½\{2n/p\} = \{n/p\}.$$

Obviously $(2n) \bmod p$ is even number (not zero), i.e. this variable can take only even values $2, 4, \ldots, p-1$. Respectively,

$$1 \le (n \bmod p) \le (p-1)/2. \tag{2.4}$$

The restriction (2.4) is necessary for other operands in (2.3). However, these operands may impose additional and more severe restrictions. Let us check it. Based on (2.4) we get the following:

$$\{(n+1)/p\} - 1/p = \{n/p\} + 1/p - 1/p = \{n/p\}.$$

Thus the equality (2.3) is proved. Note that the constraints (2.2) and (2.4) are equivalent. Recall that many prime factors (distinct or repeated) of the Catalan number are not included in the distinct-layer, since these primes belong to other layers.

Case 2.

$$\lfloor (2n-1)/p \rfloor \text{ is an odd number.} \tag{2.5}$$

Let us convert (2.1) again:

$$\begin{aligned} &½\,(\lfloor (2n-1)/p \rfloor + 1) - ((n+1)/p - \{(n+1)/p\}) \\ &\quad = \{(n+1)/p\} - 1/p - ½\,(\{(2n-1)/p\} + 1/p) + ½ \\ &\quad = \{(n+1)/p\} - 1/p + ½\,((p-1)/p - \{(2n-1)/p\}) = 0 \text{ or } 1. \end{aligned} \tag{2.6}$$

The expressions (2.3) and (2.6) differ in ½. Therefore, the last equality is ambiguous. In (2.6) the factor ½ is valid if and only if

$(2n-1) \bmod p$ is even or zero, because $p-1$ is an even number.

This is consistent with the obvious fact: the odd dividend $2n-1$ is being divided by the odd divisor p leaves an even remainder or zero if the quotient $\lfloor (2n-1)/p \rfloor$ is an odd number. So, the expression $(2n-1) \bmod p$ can take values $0, 2, 4, \ldots, p-1$.

Obviously the operand with the factor ½ take on the inverted values after subtraction, i.e. 0 goes to $p-1$, 2 goes to $p-3$, and so forth, and finally $p-1$ is inverted to 0. The result is the same set of values $0, 2, 4, \ldots, p-1$. After dividing by 2, the right operand in (2.6) can takes values $0, 1, 2 \ldots (p-1)/2$. For further analysis, connect the left operand $\{(n+1)/p\}$.

The expression (2.6) can take binary values 0 or 1. Zero means that p does not belong to the distinct-layer for a given n when the restriction (2.5) is performed. Zero is possible only in two subcases:

Subcase 2.1. $n+1 \equiv 0 \pmod{p} \Leftrightarrow 2n-1 \equiv p-3 \pmod{p}$.
Subcase 2.2. $n+1 \equiv 1 \pmod{p} \Leftrightarrow 2n-1 \equiv p-1 \pmod{p}$.

In each subcase, there are two equivalent and interrelated modular expressions (one derived from the other). In other subcases, the equation (2.6) is not zero. Let's formulate the theorem that helps us to determine the distinct-layer content.

Theorem 2.1. *Let $p>2$ be a prime. Then for the nth Catalan number, p belongs to the distinct-layer if and only if the following conditions true:*
(1) $\lfloor (2n-1)/p \rfloor$ is an odd number,
(2) $(n+1) \bmod p > 1$.

Subcases 2.1 and 2.2 clarify the distribution of 3. The expression $(2n-1) \bmod 3$ can be 0 (i.e. $p-3$) or only even 2 (i.e. $p-1$). The following consequence is obvious.

Corollary 2.1. *For any Catalan number, the prime 3 does not belong to the distinct-layer.*

3. Square-layer, cube-layer, and others

Similar to the distinct-layer, the adjacent square-layer contains the majority of squares, i.e. prime factors that have the power 2. Actually this is also true for an arbitrary layer.

Proposition 3.1. *The jth layer accumulates the bulk of prime factors that have the power j.*

In the following example, let's be clear on the figures the situation with the number of identical elements in layers.

Example 3.1. The prime factorization of the *million*th Catalan number contains 101543 primes and prime powers (including seven of 2). The number of non-repeated prime factors is equal to 101385, of which 101318 primes are distributed in the distinct-layer (in this layer, there are a total of 101382 elements). The number of prime squares is equal to 61, of which 56 are distributed in the square-layer (in the second layer, there are a total of 120 elements). The number of prime cubes (prime factors have the power 3) is equal to 6, of which 5 are distributed in the cube-layer.

The prime factorization of the *billion*th Catalan number contains 1373 prime squares, of which 1341 are distributed in the square-layer. The number of prime cubes is equal to 36, of which 30 are distributed in the cube-layer.

In (1.4-1.5) the second terms in the sums give rise to square-layer elements. Let's retain the second summands, i.e. $j=2$. Then for any prime $p>2$, we get the *square-layer relation*

$$½\left(\lfloor(2n-1)/p^2\rfloor_{\text{odd}}+1\right) - \lfloor(n+1)/p^2\rfloor = \begin{cases} 1 & \text{if } p \in L_2, \\ 0 & \text{if } p \notin L_2. \end{cases}$$

A binary result is obvious, since any prime either belongs to the square-layer or not. In the general case for *j*th layer, we have the *j*th *layer relation*

$$½\left(\lfloor(2n-1)/p^j\rfloor_{\text{odd}}+1\right) - \lfloor(n+1)/p^j\rfloor = \begin{cases} 1 & \text{if } p \in L_j, \\ 0 & \text{if } p \notin L_j. \end{cases} \qquad (3.1)$$

You can check that the above expression is equal to 1 for any prime in the main segment $(\sqrt[j]{n+1}, \sqrt[j]{2n})$. Each prime $p > \sqrt[j]{2n}$ does not belong to L_j, since both operands in (3.1) are equal to 0.

Further analysis will be done by analogy with the distinct layer (see previous section). The results obtained for the first layer are easily converted to other layers. For example, the second operand in (3.1) is calculated as follows:

$$\lfloor(n+1)/p^j\rfloor = (n+1)/p^j - \{(n+1)/p^j\}.$$

In connection with an odd floor function we also consider two cases. The value of $\lfloor(2n-1)/p^j\rfloor$ may be even or odd (zero is excluded as we take $p < \sqrt[j]{2n}$). In the first case

$$\lfloor(2n-1)/p^j\rfloor = \text{even} \qquad (3.2)$$

we get the general form of the equation (2.3):

$$\{(n+1)/p^j\} - 1/p^j - ½\left(\{(2n-1)/p^j\} + 1/p^j\right) = 0.$$

The resulting equality is easy to prove by modular arithmetic. As a result we get the analogue of equation (2.4)

$$1 \le (n \bmod p^j) \le (p^j-1)/2. \qquad (3.3)$$

It is true that both constraints (3.2) and (3.3) are equivalent. It remains to consider the second case:

$$\lfloor(2n-1)/p^j\rfloor = \text{odd}. \qquad (3.4)$$

In this variant the relation (3.1) takes the following form:

$$\{(n+1)/p^j\} - 1/p^j + ½\left((p^j-1)/p^j - \{(2n-1)/p^j\}\right) = 0 \text{ or } 1. \qquad (3.5)$$

The above expression is a general form of the previously discussed expression (2.6). Let us repeat the analysis of (2.6) with respect to an arbitrary layer.

In (3.5) the factor ½ is valid if and only if

$r = (2n-1) \bmod p^j$ is even or zero, because p^j-1 is an even number.

So, the remainder r can take values $0, 2, 4, \ldots, p^j-1$. The operand with the factor ½ take on the inverted value of r, i.e. 0 goes to p^j-1, 2 goes to p^j-3, and so forth,

and finally p^j-1 is inverted to 0. The result is an identical set of values, namely: p^j-1, p^j-3, p^j-5, …, 0. After dividing by 2, the right operand in (3.5) takes values $0, 1, 2, ..., (p^j-1)/2$. Now let's look at the first operand $\{(n+1)/\,p^j\}$ in (3.5).

The expression (3.5) can be 0 or 1. Zero means that the prime p does not belong to the jth layer for a given n. It is easy to see that zero is possible only in two ways:

(1) $n+1 \equiv 0 \pmod{p^j} \;\Leftrightarrow\; 2n-1 \equiv p^j-3 \pmod{p^j}$;
(2) $n+1 \equiv 1 \pmod{p^j} \;\Leftrightarrow\; 2n-1 \equiv p^j-1 \pmod{p^j}$.

In each row, we have two equivalent and interrelated expressions (one derived from the other). We formulate a generalized theorem for the prime identity.

Theorem 3.1. *Let $p>2$ be a prime. Then for the nth Catalan number, p belongs to the jth layer if and only if the following conditions true:*
(a) $\lfloor (2n-1)/p^j \rfloor$ is an odd number,
(b) $(n+1) \bmod p^j > 1$.

By using Theorem 3.1, you can easily choose all primes and prime powers if you browse all layers. Obviously for nth Catalan number, the total number of layers does not exceed $\log_3 n$.

4. Service online

Finally, consider the software system for the implementation of test calculation online. Let's list some programs with brief description of their functions (additional software service described in [Ep15] and [Er16]).

1. For the nth Catalan number, this program allows you to obtain all cores starting with the number $k = \lfloor \log_3 n \rfloor$ and ending with the first (main) core. In the transition from one core to another, the program displays only the additional primes and prime powers, i.e. the prime factors of the expansion in the next core.

2. The reader can obtain the layer-splitting of the prime factors of any Catalan number. Each layer contains only unique (non-repeated) primes. On the printout you can see how repetitive prime factors are divided into different layers.

3. Some programs handle only one layer of the prime factors of Catalan numbers. You can display separately distinct-layer, square-layer or cube-layer. Working with an individual layer, it is possible to obtain the information for the Catalan numbers with very large indexes (billion and more). For large indexes, the higher prime numbers are grouped in segments.

The software package uses only HTML, CSS, and JavaScript. All programs are written by me, anyone can get any texts. Participation in the further development of the service is welcomed. Again my e-mail: argenns@gmail.com.

References

[A000108] Sloane N. J. A. The On-Line Encyclopedia of Integer Sequences. https://oeis.org/A000108

[Adam05] Adamchik V. *Modular Arithmetic.* Concepts of Mathematic (2005). http://www.cs.cmu.edu/~adamchik/21-127/lectures/congruences_print.pdf

[Er16] Eremin G. Factoring a Catalan Number into Chebyshev's Segments (2016). http://eremin.xyz/catalan-cheb-2016.pdf

[Pom13] Pomerance C. *On numbers related to Catalan numbers.* Mathematics Department, Dartmouth College, Hanover (September 2013). https://math.dartmouth.edu/~carlp/catalan

[Pom15] Carl Pomerance. *Divisors of the middle binomial coefficient.* Amer. Math. Monthly **122** (2015), 636–644. https://math.dartmouth.edu/~carlp/amm2015.pdf

[Stan15] Stanley, R. P. (2015): Catalan Numbers. Cambridge (2015). http://www.cambridge.org/ro/academic/subjects/mathematics/discrete-mathematics-information-theory-and-coding/catalan-numbers?format=HB

[Wei16] Weisstein, Eric W. Catalan Numbers (Wolfram MarhWorld). http://mathworld.wolfram.com/CatalanNumber.html

[Ep15] Еремин Геннадий. Факторизация чисел Каталана. http://eremin.xyz/catalan/