

On Moessner's Theorem

Dexter Kozen
 Computer Science Department
 Cornell University
 Ithaca, NY 14853-7501, USA

Alexandra Silva*
 Centrum Wiskunde & Informatica
 Science Park 123
 1098 XG Amsterdam
 The Netherlands

June 12, 2011

Abstract

Moessner's theorem describes a procedure for generating a sequence of n integer sequences that lead unexpectedly to the sequence of n th powers $1^n, 2^n, 3^n, \dots$. *Paasche's theorem* is a generalization of Moessner's; by varying the parameters of the procedure, one can obtain the sequence of factorials $1!, 2!, 3!, \dots$ or the sequence of superfactorials $1!!, 2!!, 3!!, \dots$. *Long's theorem* generalizes Moessner's in another direction, providing a procedure to generate the sequence $a \cdot 1^{n-1}, (a+d) \cdot 2^{n-1}, (a+2d) \cdot 3^{n-1}, \dots$. Proofs of these results in the literature are typically based on combinatorics of binomial coefficients or calculational scans. In this note we give a short and revealing algebraic proof of a general theorem that contains Moessner's, Paasche's, and Long's as special cases. We also prove a generalization that gives new Moessner-type theorems.

1 Introduction

Consider the following procedure for generating $n \geq 1$ infinite sequences of positive integers. To generate the first sequence, write down the positive integers $1, 2, 3, \dots$, then cross out every n th element. For the second sequence, compute the prefix sums of the first sequence, ignoring the crossed-out elements, then cross out every $(n-1)$ st element. For the third sequence, compute the prefix sums of the second sequence, then cross out every $(n-2)$ nd element, and so on. For example, for $n = 4$,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	3	6		11	17	24		33	43	54		67	81	96		113	131	150		171	193	216	
1	4		15	32		65	108		175	256		369	500		671	864							
1			16			81			256			625				1296							

Moessner's theorem says that the final sequence is $1^n, 2^n, 3^n, \dots$.

This construction is an interesting combinatorial curiosity that has attracted much attention over the years. Moessner's theorem was never proved by its eponymous discoverer [7]. The first proof

*Until August 2011: Computer Science Department, Cornell University, Ithaca, NY 14853-7501, USA

was given later by Perron [12]. Since then, the theorem has been the subject of several popular accounts [1,2,4,6].

In the construction of Moessner's theorem, the initial step size n is constant. What happens if we increase it in each step? Let us repeat the construction starting with a step size of one and increasing the step size by one each time. Thus, in the first sequence, we cross out $1, 3, 6, 10, \dots, \binom{k+1}{2}, \dots$.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	2		6	11		18	26	35		46	58	71	85		101	118	136	155	175	
		6				24	50			96	154	225			326	444	580	735		
						24				120	274				600	1044	1624			
										120					720	1764				
															720					

Now the final sequence consists of the factorials $1, 2, 6, 24, 120, \dots = 1!, 2!, 3!, 4!, 5!, \dots$.

Let us now *increment the increment* by one in each step, thus incrementing the step size by 1, 2, 3, 4, ... in successive steps, crossing out $1, 4, 10, 20, \dots, \binom{k+2}{3}, \dots$.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	...
	2	5		10	16	23	31	40		51	63	76	80	95	...					
		2		12	28	51	82			133	196	272	352	...						
				12	40	94				224	420	692	...							
				12	52					276	696	...								
				12						288	984	...								
										288	1272	...								
										288	...									
										288	...									
										288	...									

The final sequence consists of the *superfactorials*

$$1, 2, 12, 288, \dots = 1!, 2!1!, 3!2!1!, 4!3!2!1!, \dots = 1!, 2!!, 3!!, 4!!, \dots$$

The generalization of Moessner's theorem that handles these cases is known as *Paasche's theorem* [9-11].

Long [5,6] discovered the following alternative procedure and generalization. Consider the figure illustrating the Moessner construction for $n = 4$ above. Breaking the figure into separate triangles and adding a row of 1's at the top, the first four triangles are

1	1	1	1	1		1	1	1	1	1		1	1	1	1	1		1	1	1	1	1
1	2	3	4			5	6	7	8			9	10	11	12			13	14	15	16	
1	3	6				11	17	24				33	43	54				67	81	96		
1	4					15	32					65	108					175	256			
1						16						81						256				

Call these the *level- n Moessner triangles*. The first triangle is the well-known Pascal triangle. However, note that all the triangles satisfy the *Pascal property*: each interior element is the sum of the elements immediately above it and to its left. Note also that the first column of each triangle consists of the prefix sums of the n th northeast-to-southwest row of the previous triangle. For example, the first column of the third triangle is 1, 9, 33, 65, 81, which are the prefix sums of 1, 8, 24, 32, 16, the last northeast-to-southwest row of the second triangle. Thus, to generate the next triangle in the sequence, let its first column be the prefix sums of the n th northeast-to-southwest

row of the previous triangle, let the top horizontal row consist of all 1's, and complete the triangle using the Pascal property.

Long [5,6] and Salié [13] also generalized Moessner's result to apply to the situation in which the first sequence is not the sequence of successive integers $1, 2, 3, \dots$ but the arithmetic progression $a, a + d, a + 2d, \dots$. This corresponds to a sequence of triangles with d, d, d, \dots along the top and d, a, a, a, \dots as the first column of the first triangle. They showed that the final sequence obtained by the Moessner construction is $a \cdot 1^{n-1}, (a + d) \cdot 2^{n-1}, (a + 2d) \cdot 3^{n-1}, \dots$.

Very recently, Hinze [3] and Niqui and Rutten [8] have given proofs involving concepts from functional programming, Hinze using calculational scans and Niqui and Rutten using coalgebra of streams. The proof of Hinze covers Moessner's and Paasche's result whereas Rutten and Niqui only provide a proof of the original Moessner's theorem.

The proof we present here has the advantage of covering all the theorems mentioned above and, furthermore, opening the door to new generalizations of Moessner's original result.

2 Algebraic Representation

In this section, we will describe Long's construction in terms of multidimensional generating functions. This will serve as basis for our main theorem (Theorem 2.4), which will have as corollaries Moessner's (Corollary 2.5), Long's (Corollary 2.6) and Paasche's (Corollary 2.7) results.

We represent triangles as formal power series in $\mathbb{Z}[[x, y]]$. For example, the *Pascal triangle* $\Delta = \Delta(x, y)$ is

$$\Delta(x, y) = \frac{1}{1 - (x + y)} = \sum_{d=0}^{\infty} (x + y)^d = \sum_{i, j \geq 0} \binom{i + j}{i} x^i y^j. \quad (1)$$

In this representation, the " n th northeast-to-southwest row" of $p \in \mathbb{Z}[[x, y]]$ is the homogeneous component of degree n , denoted $[p]_n$. To make this a power series in one variable x , we evaluate at $y = 1$. The operation of "taking prefix sums" is multiplying by $\sum_{i=0}^{\infty} x^i = (1 - x)^{-1}$.

To describe the operation of "completing the triangle using the Pascal property," we need some lemmas. Call a power series in x, y *Pascal* if the coefficient of every interior monomial m (a monomial of positive degree in both x and y) is the sum of the coefficients of m/x and m/y .

Lemma 2.1 $f = f(x, y)$ is *Pascal* iff

$$f = ((1 - x)f(x, 0) + (1 - y)f(0, y) - f(0, 0)) \cdot \Delta.$$

Proof. The Pascal condition says that any interior monomial has the same coefficient in f and $(x + y)f$; in other words, $(1 - x - y)f = f/\Delta$ has no interior terms. Thus

$$f/\Delta = f/\Delta|_{y=0} + f/\Delta|_{x=0} - f/\Delta|_{x=y=0} = (1 - x)f(x, 0) + (1 - y)f(0, y) - f(0, 0).$$

Multiplying by Δ gives the result. □

Lemma 2.2 If $p_x \in \mathbb{Z}[[x]]$, $p_y \in \mathbb{Z}[[y]]$, and $p_x(0) = p_y(0) = p_0$, then

$$p = ((1-x)p_x + (1-y)p_y - p_0) \cdot \Delta$$

is the unique $p \in \mathbb{Z}[[x, y]]$ such that $p(x, 0) = p_x$, $p(0, y) = p_y$, and p is Pascal. □

Proof. Immediate from Lemma 2.1. □

In particular, for $e \in \mathbb{Z}[[x]]$, $e \cdot \Delta$ is the unique $p \in \mathbb{Z}[[x, y]]$ such that $p(x, 0) = e/(1-x)$, $p(0, y) = e(0)/(1-y)$, and p is Pascal.

Now we see that each successive level- n Moessner triangle is obtained from the previous by taking the homogeneous component of degree n , evaluating at $y = 1$, and multiplying by Δ . In other words, if we define inductively

$$h_0(x, y) = 1 \qquad h_{k+1}(x, y) = [h_k(x, 1) \cdot \Delta(x, y)]_n,$$

then the k th level- n Moessner triangle is $h_k(x, 1) \cdot \Delta$ and the final sequence in the Moessner construction is the lead coefficient of $h_k(x, 1)$ for $k = 1, 2, 3, \dots$

More generally, let $h_0 \in \mathbb{Z}[[x, y]]$ be an arbitrary homogeneous polynomial, and let d_0, d_1, d_2, \dots be an arbitrary sequence of nonnegative integers. Define inductively

$$h_{k+1}(x, y) = [h_k(x, 1) \cdot \Delta(x, y)]_{\deg h_k + d_k}. \tag{2}$$

The Moessner construction is the special case $h_0 = 1$, $d_0 = n$, and $d_i = 0$ for $i \geq 1$.

Lemma 2.3 Let $h(x, y)$ be homogeneous of degree m and let $d \geq 0$. Then

$$[h(x, 1) \cdot \Delta(x, y)]_{m+d} = (x+y)^d h(x, x+y).$$

Proof. Summing both sides over all $d \geq 0$ and using (1), it suffices to show that any monomial of degree m or greater has the same coefficient in the two power series

$$h(x, 1) \cdot (1 - (x+y))^{-1} \qquad h(x, x+y) \cdot (1 - (x+y))^{-1}.$$

In other words, all terms of degree m or greater in $(h(x, 1) - h(x, x+y)) \cdot (1 - (x+y))^{-1}$ vanish. But if $h(x, y) = \sum_{i+j=m} a_{ij} x^i y^j$, then

$$(h(x, 1) - h(x, x+y)) \cdot (1 - (x+y))^{-1} = \sum_{i+j=m} a_{ij} x^i \left(\frac{1 - (x+y)^j}{1 - (x+y)} \right) = \sum_{i+j=m} a_{ij} x^i \left(\sum_{k=0}^{j-1} (x+y)^k \right),$$

a polynomial of degree at most $m - 1$. □

Theorem 2.4 Let h_k be the sequence defined by (2). For all $k \geq 0$,

$$h_k(x, y) = \prod_{i=0}^{k-1} ((k-i)x + y)^{d_i} \cdot h_0(x, kx + y).$$

Proof. By induction on k . The basis $k = 0$ is trivial. For the induction step,

$$\begin{aligned}
h_{k+1} &= [h_k(x, 1) \cdot \Delta(x, y)]_{\deg h_k + d_k} && \text{by (2)} \\
&= (x + y)^{d_k} h_k(x, x + y) && \text{by Lemma 2.3} \\
&= (x + y)^{d_k} \prod_{i=0}^{k-1} ((k - i)x + x + y)^{d_i} h_0(x, kx + x + y) && \text{by the induction hypothesis} \\
&= \prod_{i=0}^k ((k + 1 - i)x + y)^{d_i} h_0(x, (k + 1)x + y) && \text{by simplification.}
\end{aligned}$$

□

Paasche's, Long's, and Moessner's theorems are now immediate consequences of Theorem 2.4.

Corollary 2.5 (Moessner's Theorem) *If $h_0 = 1$, $d_0 = n$, and $d_k = 0$ for $k \geq 1$, then the lead coefficient of $h_k(x, 1)$ is k^n for all $k \geq 1$.*

Proof. Immediate from Theorem 2.4 by substituting the given values and simplifying. □

Corollary 2.6 (Long's Theorem) *If $h_0 = (a - d)x + dy$, $d_0 = n - 1$, and $d_k = 0$ for $k \geq 1$, then the lead coefficient of $h_k(x, 1)$ is $(a + (k - 1)d)k^{n-1}$ for all $k \geq 1$.*

Proof. Immediate from Theorem 2.4 by substituting the given values and simplifying. The definition of h_0 sets the first column of the first triangle to d, a, a, a, \dots . □

Corollary 2.7 (Paasche's Theorem) *For $h_0 = 1$ and any sequence d_0, d_1, d_2, \dots , the lead coefficient of $h_k(x, 1)$ is*

$$\prod_{i=0}^{k-1} (k - i)^{d_i} \tag{3}$$

for all $k \geq 0$. In particular, the sequences $d = 1, 1, 1, \dots$ and $d = 1, 2, 3, \dots$ yield the factorials and superfactorials, respectively.

Proof. The expression (3) is immediate from Theorem 2.4. For the sequence $d = 1, 1, 1, \dots$, the lead coefficient of $h_k(x, 1)$ is

$$\prod_{i=0}^{k-1} (k - i)^{d_i} = \prod_{i=0}^{k-1} (k - i) = k!$$

For the sequence $d = 1, 2, 3, \dots$, we happily calculate:

$$\prod_{i=0}^{k-1} (k - i)^{d_i} = \prod_{i=0}^{k-1} (k - i)^{i+1} = k(k - 1)^2(k - 2)^3 \dots 1^k = k!(k - 1)! \dots 1! = k!!$$

In general, $d_i = \binom{i+m}{m}$ gives $k \underbrace{!! \dots !}_{m+1}$, where $k \underbrace{!! \dots !}_{m+1} = \prod_{i=0}^{k-1} (k - i) \underbrace{!! \dots !}_m$. □

3 A Multidimensional Moessner Theorem

Lemma 2.3 and Theorem 2.4 have the following multidimensional generalization.

Let $h(x_1, \dots, x_k) \in \mathbb{Z}[x_1, \dots, x_k]$ be a homogeneous polynomial of degree m . Let $w = w_1, \dots, w_k \in \mathbb{Z}[x_1, \dots, x_n]$ be a sequence of k homogeneous linear forms in n variables. Let $d \geq 0$.

We write $h(w_1, \dots, w_k)$ for the polynomial obtained by substituting w_i for x_i , $1 \leq i \leq k$. We can regard h as a polynomial function of abstract type $R^k \rightarrow R$ and w as a linear map of abstract type $R^n \rightarrow R^k$. In this view, $h(w_1, \dots, w_k)$ represents the functional composition $h \circ w : R^n \rightarrow R$. The map w can be represented by a $k \times n$ matrix whose ij th entry is the coefficient of x_j in w_i .

Lemma 2.3 can then be generalized as follows.

Lemma 3.1 *Let $h \in \mathbb{Z}[x_1, \dots, x_k]$ be homogeneous of degree m . Let $w = w_1, \dots, w_k \in \mathbb{Z}[x_1, \dots, x_n]$ be homogeneous linear forms. For any $d \geq 0$,*

$$\left[h(w_1, \dots, w_{k-1}, 1) \cdot (1 - w_k)^{-1} \right]_{m+d} = w_k^d \cdot h(w_1, \dots, w_k).$$

Proof. The proof is the same as Lemma 2.3. Summing over all $d \geq 0$ and using the fact that $\sum_{d=0}^{\infty} w_k^d = (1 - w_k)^{-1}$, the lemma says that all terms of degree m or greater in

$$(h(w_1, \dots, w_{k-1}, 1) - h(w_1, \dots, w_k)) \cdot (1 - w_k)^{-1}$$

vanish. But if $a_{i_1 \dots i_k}$ is the coefficient of $x_1^{i_1} \cdots x_k^{i_k}$ in h ,

$$\begin{aligned} & (h(w_1, \dots, w_{k-1}, 1) - h(w_1, \dots, w_k)) \cdot (1 - w_k)^{-1} \\ &= \sum_{i_1 + \dots + i_k = m} a_{i_1 \dots i_k} w_1^{i_1} \cdots w_{k-1}^{i_{k-1}} \left(\frac{1 - w_k^{i_k}}{1 - w_k} \right) \\ &= \sum_{i_1 + \dots + i_k = m} a_{i_1 \dots i_k} w_1^{i_1} \cdots w_{k-1}^{i_{k-1}} \left(\sum_{i=0}^{i_k-1} w_k^i \right), \end{aligned}$$

a polynomial of degree at most $m - 1$. □

Lemma 3.1 describes a map

$$T_{w,d}(h) = \left[h(w_1, \dots, w_{k-1}, 1) \cdot (1 - w_k)^{-1} \right]_{m+d} = w_k^d \cdot h(w_1, \dots, w_k)$$

parameterized by w and d from a homogeneous polynomial h of degree m to a new homogeneous polynomial $w_k^d \cdot h \circ w$ of degree $m + d$.

A generalization of Theorem 2.4 can now be stated. Let $m_k \geq 0$ and $d_k \geq 0$ for $k \geq 0$. Let $h_0 \in \mathbb{Z}[x_1, \dots, x_{m_0}]$ be a homogeneous polynomial, considered as a polynomial function of abstract type $R^{m_0} \rightarrow R$. Let $w(k) : R^{m_{k+1}} \rightarrow R^{m_k}$ be a sequence of linear maps. Define

$$w(i; k) = w(i) \circ w(i+1) \circ \cdots \circ w(k) : R^{m_{k+1}} \rightarrow R^{m_i}.$$

Then $w(i; k)$ is a linear map whose matrix representation is the product of the matrices representing $w(i), w(i+1), \dots, w(k)$ in that order. By convention, we take $w(k+1; k)$ to be the identity map $R^{m_k} \rightarrow R^{m_k}$. Define inductively

$$h_{k+1} = T_{w(k), d_k}(h_k) = w(k)_{m_k}^{d_k} \cdot (h_k \circ w(k)) : R^{m_{k+1}} \rightarrow R.$$

It follows inductively that h_k is homogeneous of degree $\deg h_0 + \sum_{i=0}^{k-1} d_i$ in m_k variables.

Theorem 3.2

$$h_k = \prod_{i=0}^{k-1} (w(i)_{m_i}^{d_i} \circ w(i+1; k-1)) \cdot (h_0 \circ w(0; k-1)).$$

Proof. We proceed by induction on k . The basis $k = 0$ is trivial. For the induction step,

$$\begin{aligned} h_{k+1} &= T_{w(k), d_k}(h_k) \\ &= w(k)_{m_k}^{d_k} \cdot (h_k \circ w(k)) \\ &= w(k)_{m_k}^{d_k} \cdot \left(\prod_{i=0}^{k-1} (w(i)_{m_i}^{d_i} \circ w(i+1; k-1)) \cdot (h_0 \circ w(0; k-1)) \right) \circ w(k) \\ &= (w(k)_{m_k}^{d_k} \cdot w(k+1; k)) \cdot \prod_{i=0}^{k-1} (w(i)_{m_i}^{d_i} \circ w(i+1; k)) \cdot (h_0 \circ w(0; k)) \\ &= \prod_{i=0}^k (w(i)_{m_i}^{d_i} \circ w(i+1; k)) \cdot (h_0 \circ w(0; k)). \end{aligned}$$

The reasoning is the same as in the proof of Theorem 2.4. □

This generalization leads to many other Moessner-type theorems. We can obtain any order- m linear recurrence with $h_0 = 1$, $d_0 = 1$ and $d_k = 0$ for $k \geq 1$, and $m_k = m$ for $k \geq 0$. In this case Theorem 3.2 gives

$$h_0 = 1 \qquad h_k = w_m \circ w^{(k-1)}$$

where $w^{(i)}$ denotes the i th compositional power of w . For example, to obtain the Fibonacci sequence, we can take w to be the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

corresponding to $w_1 = x + y$ and $w_2 = x$.

4 Relation to the Pascal Simplex

The original Moessner–Long construction has a geometric interpretation that can be explained in terms of the k -dimensional Pascal simplex

$$\Delta_k(x_1, \dots, x_k) = \sum_{n=0}^{\infty} (x_1 + \dots + x_k)^n = \frac{1}{1 - (x_1 + \dots + x_k)}.$$

Define $h_k : R^{k+1} \rightarrow R$ inductively:

$$h_0 = x_1^n \quad h_{k+1}(x_1, \dots, x_{k+2}) = [h_k(x_1, \dots, x_k, 1) \cdot \Delta_2(x_{k+1}, x_{k+2})]_n.$$

Theorem 3.2 gives

$$h_k(x_1, \dots, x_{k+1}) = (x_1 + \dots + x_{k+1})^n,$$

the n th homogeneous component of Δ_{k+1} . The sequence of triangles in the Moessner–Long construction is obtained by collapsing the multidimensional simplices in this construction back to dimension two in each step by substituting 1 for one of the variables. The final Moessner sequence $1^n, 2^n, 3^n, \dots$ is the same as the sequence obtained by summing the coefficients of the degree- n homogeneous component of the Pascal simplex of dimension $1, 2, 3, \dots$:

$$(x_1 + \dots + x_k)^n \Big|_{x_1=\dots=x_k=1} = k^n.$$

The only difference is that instead of collapsing back to dimension two in each step, we build up a multidimensional simplex and collapse only at the end.

References

- [1] J. H. Conway and R. K. Guy. Moessner’s magic. In *The Book of Numbers*, pages 63–65. Springer-Verlag, 1996.
- [2] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics*. Addison Wesley, 2nd edition, 1994.
- [3] R. Hinze. Scans and convolutions—a calculational proof of Moessner’s theorem. In Sven-Bodo Scholz, editor, *Post-proceedings of the 20th International Symposium on the Implementation and Application of Functional Languages (IFL ’08)*, volume 5836 of *Lecture Notes in Computer Science*. Springer-Verlag, 2009.
- [4] R. Honsberger. *More Mathematical Morsels*. Math. Assoc. Amer., 1991.
- [5] C. T. Long. On the Moessner theorem on integral powers. *Amer. Math. Monthly*, 73(8):846–851, October 1966.
- [6] C. T. Long. Strike it out—add it up. *The Mathematical Gazette*, 66(438):273–277, December 1982.
- [7] A. Moessner. Eine Bemerkung über die Potenzen der natürlichen Zahlen. *Sitzungsberichten der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse 1952*, 29, March 1951.
- [8] M. Niqui and J. Rutten. An exercise in coinduction: Moessner’s theorem. Technical Report SEN-1103, CWI Amsterdam, 2011.
- [9] I. Paasche. Ein neuer Beweis des moessnerischen Satzes. *Sitzungsberichten der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse 1952*, 1:1–5, February 1953.
- [10] I. Paasche. Ein zahlentheoretische-logarithmischer Rechenstab. *Math. Naturwiss. Unterr.*, 6:26–28, 1953–54.
- [11] I. Paasche. Eine Verallgemeinerung des moessnerschen Satzes. *Compositio Mathematica*, 12:263–270, 1954.
- [12] O. Perron. Beweis des Moessnerschen Satzes. *Sitzungsberichten der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse 1951*, 4:31–34, May 1951.
- [13] H. Salié. Bemerkung zu einem Satz von Moessner. *Sitzungsberichten der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse 1952*, 2:7–11, February 1952.