Counting Abelian Squares

L. B. Richmond

Department of Combinatorics and Optimization
University of Waterloo
Waterloo, ON N2L 3G1
Canada

`lbrichmo@math.uwaterloo.ca`

Jeffrey Shallit

School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada

`shallit@cs.uwaterloo.ca`

Submitted: Jan 5, 2009; Accepted: Jun 8, 2009; Published: Jun 19, 2009
Mathematics Subject Classification: 68R15, 05A16

Abstract

An *abelian square* is a nonempty string of length $2n$ where the last n symbols form a permutation of the first n symbols. Similarly, an *abelian r'th power* is a concatenation of r blocks, each of length n, where each block is a permutation of the first n symbols. In this note we point out that some familiar combinatorial identities can be interpreted in terms of abelian powers. We count the number of abelian squares and give an asymptotic estimate of this quantity.

1 Introduction

An *abelian square* of length $2n$ is a nonempty string of the form xx', where $|x| = |x'| = n > 0$ and x' is a permutation of x. Two abelian squares in English are `reappear` and `intestines`. Of course, the permutation can be the identity, so ordinary squares such as `murmur` and `hotshots` are also considered to be abelian squares. Similarly, an *abelian r'th power* is a concatenation of r blocks, each of length n, where each block is a permutation of the first n symbols. For example, `deeded` is an abelian cube.

Abelian squares were introduced by Erdős [10, p. 240] and since then have been extensively studied in the combinatorics on words literature (see, for example, [1, p. 37]).

In this note we point out that some familiar combinatorial identities can be interpreted in terms of counting abelian powers. We discuss enumerating the abelian squares over an alphabet of size k and give an asymptotic estimate for this quantity.

2 Preliminaries

Let $f_k(n)$ be the number of abelian squares of length $2n$ over an alphabet Σ with k letters. Without loss of generality, we assume that $\Sigma = \{1, 2, \ldots, k\}$.

Given a string x with $|x| = n$, the signature of x is defined to be the vector enumerating the number of 1's, 2's, etc. in x. (In computer science, this vector is sometimes called the Parikh vector.) For example, the signature of 213313 is $(2, 1, 3)$. Hence a string xx' is an abelian square iff the signatures of x and x' are the same.

The following table enumerates $f_k(n)$ for the first few values of k and n, together with the sequence numbers from Sloane's *Encyclopedia* [18].

$k\backslash n$	0	1	2	3	4	5	6	Sloane
1	1	1	1	1	1	1	1	A000012
2	1	2	6	20	70	252	924	A000984
3	1	3	15	93	639	4653	35169	A002893
4	1	4	28	256	2716	31504	387136	A002895
5	1	5	45	545	7885	127905	2241225	
6	1	6	66	996	18306	384156	8848236	
Sloane	A000012	A000027	A000384					

Examination of this table suggests that $f_2(n) = \binom{2n}{n}$, and indeed, this can be proved as follows. Suppose we choose the positions of the 1's in the first n symbols; if there are i of them, this can be done in $\binom{n}{i}$ ways. Once we choose these, the remaining symbols of the first n must be 2's. The last n symbols must have the same signature as the first n, and this can be done in $\binom{n}{i}$ ways. So we get

$$f_2(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 .$$

The sequence $f_2(n)$ is sequence A000984 in Sloane's *On-line Encyclopedia of Integer Sequences* [18].

There is a nice combinatorial proof that this sum is actually $\binom{2n}{n}$. Consider a string of length $2n$, and choose n positions in it. If a position falls in the first half of the string, make it 1; if a position falls in the last half of the string, make it 2. Of the remaining unchosen positions, make them 2 if they fall in the first half and 1 if they fall in the last half. It is easy to see that this gives a bijection with the set of abelian squares. Thus we obtain $f_2(n) = \binom{2n}{n}$.

We can now use this idea to evaluate $f_k(n)$ in terms of $f_{k-1}(n)$. Choose the positions of the 1's in the first and last halves of the string; this can be done in $\binom{n}{i}^2$ ways. Now fill in the remaining $n-2i$ positions with $k-1$ symbols in $f_{k-1}(n-i)$ ways. Thus

$$f_k(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 f_{k-1}(n-i) = \sum_{0 \le i \le n} \binom{n}{n-i}^2 f_{k-1}(n-i) = \sum_{0 \le j \le n} \binom{n}{j}^2 f_{k-1}(j).$$

For $k = 3$ this gives

$$f_3(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 \binom{2i}{i}.$$

The sequence $f_3(n)$ is sequence A002893 in Sloane's *On-line Encyclopedia of Integer Sequences.*

More generally, we can write $f_{k_1+k_2}(n)$ in terms of $f_{k_1}(n)$ and $f_{k_2}(n)$. We have

$$f_{k_1+k_2}(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 f_{k_1}(i) f_{k_2}(n-i), \tag{1}$$

a formula originally given by Barrucand [3, 4, 17]. We can prove the formula by counting abelian squares of length kn over an alphabet of size $k = k_1 + k_2$, in two different ways.

To see this, suppose the first n symbols have i occurrences of the symbols $1, 2, \ldots, k_1$. Note that we can choose the positions where the symbols $1, \ldots, k_1$ will go in the first n symbols in $\binom{n}{i}$ ways, and where they will go in the last n symbols in $\binom{n}{i}$ ways. Once the positions are chosen, we can fill them in with $1, \ldots, k_1$ in $f_{k_1}(i)$ ways. The remaining positions can be filled with the remaining symbols $k_1+1, k_1+2, \ldots, k_1+k_2$ in $f_{k_2}(n-i)$ ways.

For $k_1 = k_2 = 2$, we get

$$f_4(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 \binom{2i}{i} \binom{2n-2i}{n-i}.$$

The sequence $f_4(n)$ is sequence A002895 in Sloane's *On-line Encyclopedia of Integer Sequences.*

A general formula is

$$f_k(n) = \sum_{n_1 + \cdots + n_k = n} \binom{n}{n_1\ n_2 \cdots\ n_k}^2, \tag{2}$$

which follows from choosing the signature of the first half of the string and then matching it in the second. Here n_i counts the number of occurrences of i, and $\binom{n}{n_1\ n_2\ \cdots\ n_k}$ is the multinomial coefficient $\frac{n!}{n_1! n_2! \cdots n_k!}$.

For $k = 3$, the formula (2) was studied by Barrucand [5, 6]; also see the paper of Callan [7]. More recently, Callan [8] has given some beautiful combinatorial interpretations that can be viewed in terms of abelian powers.

Finding a closed form for the sum in (2) was stated as a problem by Richards and Cambanis [15]. Our Theorem 4 was first conjectured (with a typographical error) by Ruehr [17]. The first author and Rousseau [16] gave a derivation of this formula based on work of Barrucand [3] and Hayman [12]. In this paper we give another derivation of this formula. Cioabă [9] mentioned the sum in (2) and say that "obtaining a closed formula ... seems to be an interesting and difficult combinatorial problem in itself".

3 Asymptotics

In this section we use the formula (2) to obtain the asymptotic behavior of $f_k(n)$ as $n \to \infty$. In what follows we shamelessly apply the factorial function to noninteger arguments, using the standard definition $x! = \Gamma(x+1)$, where Γ is the well-known gamma function.

First, let's consider the asymptotics of

$$\binom{n}{n_1\ n_2\ \cdots\ n_k}. \tag{3}$$

We use an idea that is due (more or less) to Lagrange [13]. The maximum of the multinomial coefficient (3) occurs when $n_i = \frac{n}{k}$, so write $n_i = \frac{n}{k} + x_i\sqrt{n}$. Thus

$$n = \sum_{1\le i\le k} n_i = n + \sum_{1\le i\le k} x_i\sqrt{n},$$

and so $\sum_{1\le i\le k} x_i = 0$.

Stirling's formula states that

$$n! = e^{n\log n - n}\sqrt{2\pi n}\left(1 + O(n^{-1})\right) \quad \text{as } n \to \infty. \tag{4}$$

Using Taylor's formula

$$\log(1+y) = y - \frac{y^2}{2} + O(y^3) \quad \text{with } y := \frac{x_i k}{\sqrt{n}}, \tag{5}$$

we get

$$\begin{aligned}
\log n_i &= \log\left(\frac{n}{k} + x_i\sqrt{n}\right)\\
&= \log\left(\frac{n}{k}\left(1 + \frac{x_i k}{\sqrt{n}}\right)\right)\\
&= \log\frac{n}{k} + \log\left(1 + \frac{x_i k}{\sqrt{n}}\right)\\
&= \log\frac{n}{k} + \frac{x_i k}{\sqrt{n}} - \frac{1}{2}\frac{x_i^2 k^2}{n} + O(x_i^3 n^{-3/2}).
\end{aligned}$$

Hence

$$\begin{aligned} n_i \log n_i &= \left(\frac{n}{k} + x_i\sqrt{n}\right)\left(\log\frac{n}{k} + \frac{x_i k}{\sqrt{n}} - \frac{1}{2}\frac{x_i^2 k^2}{n} + O(x_i^3 n^{-3/2})\right) \\ &= \left(\frac{n}{k} + x_i\sqrt{n}\right)\log\frac{n}{k} + \sqrt{n}x_i + \frac{1}{2}k x_i^2 + O(x_i^3 n^{-1/2}). \end{aligned}$$

Thus,

$$n_i \log n_i - n_i = \left(\frac{n}{k} + x_i\sqrt{n}\right)\log\frac{n}{k} + \frac{1}{2}k x_i^2 - \frac{n}{k} + O(x_i^3 n^{-1/2}) \tag{6}$$

and hence if $|x_i| \le n^{\epsilon}$ for some $0 < \epsilon < \frac{1}{6}$, we get

$$\sum_{1\le i\le k} (n_i \log n_i - n_i) = n\log\frac{n}{k} - n + \left(\frac{1}{2}k\sum_{1\le i\le k} x_i^2\right) + O(n^{-1/2+3\epsilon}), \tag{7}$$

where we have used the fact that $\sum_{1\le i\le k} x_i = 0$.

Thus

$$\prod_{1\le i\le k}\left(\frac{n}{k} + x_i\sqrt{n}\right)! \sim \exp\left(n\log\frac{n}{k} - n + \left(\frac{1}{2}k\sum_{1\le i\le k} x_i^2\right) + O(n^{-1/2+3\epsilon})\right)\left(2\pi\frac{n}{k}\right)^{k/2}. \tag{8}$$

Hence for $|x_i| \le n^{\epsilon}$ we get

$$\begin{aligned} \binom{n}{n_1\ n_2\ \cdots\ n_k} &= \frac{n!}{\prod_{1\le i\le k}(\frac{n}{k} + x_i\sqrt{n})!} \\ &\sim \exp\left(n\log k - \frac{k}{2}\sum_{1\le i\le k} x_i^2\right)(2\pi n)^{(1-k)/2}k^{k/2} \\ &= k^n \exp\left(-\frac{k}{2}\sum_{1\le i\le k} x_i^2\right)(2\pi n)^{(1-k)/2}k^{k/2}, \end{aligned}$$

and hence

$$\binom{n}{n_1\ n_2\ \cdots\ n_k}^2 \sim k^{2n}\exp\left(-k\sum_{1\le i\le k} x_i^2\right)(2\pi n)^{1-k}k^k. \tag{9}$$

Now let's approximate the sum

$$\sum_{n_1+n_2+\cdots+n_k=n}\binom{n}{n_1\ n_2\cdots\ n_k}^2$$

with the multiple integral

$$k^{2n}(2\pi n)^{1-k}k^k \underbrace{\int_0^n\int_0^n\cdots\int_0^n}_{k-1}\exp\left(-k\sum_{1\le i\le k} x_i^2\right)dn_1 dn_2\cdots dn_{k-1} =$$

$$k^{2n}(2\pi n)^{1-k}k^k n^{(k-1)/2} \quad \times$$

$$\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{k-1}\exp\left(-k\sum_{1\le i\le k-1}x_i^2-k\left(\sum_{1\le i\le k-1}x_i\right)^2\right)dx_1dx_2\cdots dx_{k-1}. \quad (10)$$

where we have used the fact that $dn_i = \sqrt{n}\,dx_i$ and $x_k = -x_1 - x_2 - \cdots - x_{k-1}$.

Note that the integrand is guaranteed to be asymptotic to the quantity we want only if $|x_i| \le n^\epsilon$, but outside this region the integrand is exponentially small.

In order to evaluate the multiple integral (10), we need three lemmas.

Lemma 1. *If $a > 0$, then*

$$\int_{-\infty}^{\infty}\exp\left(-(ax^2+bx+c)\right)dx = \exp\left(\frac{b^2}{4a}-c\right)\pi^{1/2}a^{-1/2}.$$

Proof. This can essentially be found, for example, in [11, Eq. 3.323.2], but for completeness we give the proof (also see [14]).

Complete the square, writing

$$ax^2+bx+c = a\left(x+\frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}.$$

Make the substitution $u = x + \frac{b}{2a}$ to get

$$\int_{-\infty}^{\infty}\exp\left(-(ax^2+bx+c)\right)dx = \exp\left(\frac{b^2}{4a}-c\right)\int_{-\infty}^{\infty}\exp(-au^2)du.$$

Now make the substitution $v = a^{1/2}u$ to get

$$\int_{-\infty}^{\infty}\exp(-au^2)du = a^{-1/2}\int_{-\infty}^{\infty}\exp(-v^2)dv.$$

The result now follows from the well-known evaluation $\int_{-\infty}^{\infty}\exp(-v^2)dv = \pi^{1/2}$. □

Lemma 2. *Let $S_{m,0} = \left(\sum_{1\le i\le m}x_i^2\right) + \left(\sum_{1\le i\le m}x_i\right)^2$, and for $1 \le l \le m$ define $S_{m,l}$ by*

$$\pi^{1/2}\left(\frac{l}{l+1}\right)^{1/2}\exp(-S_{m,l}) = \int_{-\infty}^{\infty}\exp(-S_{m,l-1})dx_l. \quad (11)$$

Then

$$S_{m,l} = \frac{l+2}{l+1}\sum_{l+1\le j\le m}x_j^2 + \frac{2}{l+1}\sum_{l+1\le i<j\le m}x_ix_j.$$

Proof. By induction on l. Clearly the result is true for $l = 0$. Now apply Lemma 1, with $a = \frac{l+2}{l+1}$, $b = \frac{2}{l+1}\sum_{l+2\le j\le m} x_j$, and $c = \frac{l+2}{l+1}\sum_{l+2\le j\le m} x_j^2 + \frac{2}{l+1}\sum_{l+2\le i<j\le m} x_i x_j$. We now have

$$\begin{aligned}
c - \frac{b^2}{4a} &= \frac{l+2}{l+1}\sum_{l+2\le j\le m} x_j^2 + \frac{2}{l+1}\sum_{l+2\le i<j\le m} x_i x_j - \frac{\frac{4}{(l+1)^2}\left(\sum_{l+2\le j\le m} x_j\right)^2}{4\frac{l+2}{l+1}} \\
&= \frac{l+2}{l+1}\sum_{l+2\le j\le m} x_j^2 + \frac{2}{l+1}\sum_{l+2\le i<j\le m} x_i x_j - \frac{\sum_{l+2\le j\le m} x_j^2}{(l+1)(l+2)} \\
&\qquad - 2\frac{\sum_{l+2\le i<j\le m} x_i x_j}{(l+1)(l+2)} \\
&= \frac{(l+2)^2-1}{(l+1)(l+2)}\sum_{l+2\le j\le m} x_j^2 + \frac{2(l+2)-2}{(l+1)(l+2)}\sum_{l+2\le i<j\le m} x_i x_j \\
&= \frac{l+3}{l+2}\sum_{l+2\le j\le m} x_j^2 + \frac{2}{l+2}\sum_{l+2\le i<j\le n} x_i x_j \\
&= S_{m,l+1}.
\end{aligned}$$

□

Thus we get

Lemma 3.

$$\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{m} \exp\left(-S_{m,0}\right) dx_1 dx_2 \cdots dx_m = \pi^{m/2}(m+1)^{-1/2}.$$

Proof. Apply Lemma 2 iteratively, obtaining

$$\begin{aligned}
&\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{m} \exp(-S_{m,0}) dx_1 dx_2 \cdots dx_m \\
&\qquad = \pi^{1/2}\left(\frac{1}{2}\right)^{1/2}\pi^{1/2}\left(\frac{2}{3}\right)^{1/2}\cdots\ \pi^{1/2}\left(\frac{m}{m+1}\right)^{1/2} = \pi^{m/2}(m+1)^{-1/2},
\end{aligned}$$

where we have used telescoping cancellation. □

It now follows (by a change of variables), that

$$\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{k-1} \exp\left(-kS_{k-1,0}\right) dx_1 dx_2 \cdots dx_{k-1} = \pi^{(k-1)/2}k^{-k/2}, \tag{12}$$

and so

$$\sum_{n_1+n_2+\cdots+n_k=n} \binom{n}{n_1\ n_2 \cdots\ n_k}^2 \sim k^{2n}(2\pi n)^{1-k} k^k n^{(k-1)/2} k^{-k/2} \pi^{(k-1)/2}$$
$$= k^{2n+k/2} 2^{1-k} \pi^{(1-k)/2} n^{(1-k)/2}.$$

We have proved

Theorem 4. *Let k be an integer ≥ 2. Then, as $n \to \infty$, we have*

$$f_k(n) \sim k^{2n+k/2}(4\pi n)^{(1-k)/2}.$$

4 Remark

Our original motivation for estimating the number of abelian squares of length $2n$ over an alphabet of size k was an attempt to use the Lovász local lemma [2, Chap. 5] to prove the existence of an infinite word avoiding abelian squares. However, since by Theorem 4 the chance that a randomly chosen string of length $2n$ is an abelian square is asymptotically

$$f_k(n)/k^{2n} \sim k^{k/2}(4\pi n)^{(1-k)/2} = \Theta(n^{(1-k)/2}),$$

this approach seems unlikely to work.

Acknowledgments

We acknowledge with thanks conversations with George Labahn, David Callan, and Stephen New. We also thank the referee for several suggestions.

References

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

[2] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley, 2000.

[3] P. Barrucand. Sur la somme des puissances des coefficients multinomiaux et les puissances successives d'une fonction de Bessel. *C. R. Acad. Sci. Paris* **253** (1964), 5318–5320.

[4] P. Barrucand. Quelques intégrales relatives aux fonctions de Bessel et aux sommes de carré des coefficients multinomiaux. *C. R. Acad. Sci. Paris* **260** (1965), 5439–5441.

[5] P. Barrucand. Problem 75-4: A combinatorial identity. *SIAM Review* **17** (1975), 168.

[6] D. R. Breach, D. McCarthy, D. Monk, and P. E. O'Neil. Comment on problem 75-4. *SIAM Review* **18** (1976), 303–304.

[7] D. Callan. A combinatorial interpretation for an identity of Barrucand. *J. Integer Sequences* **11** (2008), 08.3.4 (electronic), `http://www.cs.uwaterloo.ca/journals/JIS/VOL11/Callan2/callan204.html`

[8] D. Callan. Card deals, lattice paths, abelian words and combinatorial identities. Preprint, `http://arxiv.org/abs/0812.4784`, 2008.

[9] S. M. Cioabă. Closed walks and eigenvalues of the Abelian Cayley graphs. *C. R. Acad. Sci. Paris Ser. I* **342** (2006), 635–638.

[10] P. Erdős. Some unsolved problems. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 221–254.

[11] I. S. Gradshteyn and I. W. Ryzhik. *Tables of Integrals, Series, and Products.* Academic Press, 1965.

[12] W. K. Hayman. A generalisation of Stirling's formula. *J. reine Angew. Math.* **196** (1956), 67–95.

[13] J. L. Lagrange. Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations. *Miscellanea Taurinensia* **5** (1770–1773). Reprinted in *Oeuvres*, Vol. 2, pp. 173–234.

[14] V. S. Moll. The integrals in Gradshteyn and Rhyzik [sic]. Part 13: Evaluation using the error function. Available at `http://www.math.tulane.edu/~vhm/web_html/erfweb.pdf`, October 4 2006.

[15] D. Richards and S. Cambanis. Problem 87-2: a multinomial summation. *SIAM Review* **29** (1987), 121–122.

[16] B. Richmond and C. Rousseau. Comment on problem 87-2. *SIAM Review* **31** (1989), 122–125.

[17] O. G. Ruehr, G. E. Andrews, and L. W. Kolitsch. Comment on problem 87-2. *SIAM Review* **30** (1988), 128–130.

[18] N. J. A. Sloane. The on-line encyclopedia of integer sequences, 2008. Available at `http://www.research.att.com/~njas/sequences/` .