involve

a journal of mathematics

mathematical sciences publishers

2010 vol. 3, no. 2

involve

pjm.math.berkeley.edu/involve

PRODUCTION

Silvio Levy, Scientific Editor   Sheila Newbery, Senior Production Editor   Cover design: ©2008 Alex Scorpan

See inside back cover or http://pjm.math.berkeley.edu/involve for submission instructions.

The subscription price for 2010 is US $100/year for the electronic version, and $120/year (+$20 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94704-3840, USA.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW™ from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

http://www.mathscipub.org

A NON-PROFIT CORPORATION

Typeset in LaTeX

Recursive sequences and polynomial congruences

J. Larry Lehman and Christopher Triola

(Communicated by Kenneth S. Berenhaut)

We consider the periodicity of recursive sequences defined by linear homogeneous recurrence relations of arbitrary order, when they are reduced modulo a positive integer m. We show that the period of such a sequence with characteristic polynomial f can be expressed in terms of the order of $\omega = x + \langle f \rangle$ as a unit in the quotient ring $\mathbb{Z}_m[\omega] = \mathbb{Z}_m[x]/\langle f \rangle$. When $m = p$ is prime, this order can be described in terms of the factorization of f in the polynomial ring $\mathbb{Z}_p[x]$. We use this connection to develop efficient algorithms for determining the factorization types of monic polynomials of degree $k \leq 5$ in $\mathbb{Z}_p[x]$.

1. Introduction

This article grew out of an undergraduate research project, performed by the second author under the direction of the first, to determine if results about the periodicity of second-order linear homogeneous recurrence relations modulo positive integers could be extended to higher orders. We arrived, somewhat unexpectedly, at algorithms to determine the degrees of the irreducible factors of quintic and smaller degree polynomials modulo prime numbers. The algebraic properties of certain finite rings, particularly automorphisms of those rings, provided the connection between these two topics.

To illustrate some of the ideas in this article, we begin with the famous example of the Fibonacci sequence, defined by $F_n = F_{n-1} + F_{n-2}$ with $F_0 = 0$ and $F_1 = 1$. If, for some positive integer m, we replace each F_n by its remainder on division by m, we obtain a new sequence of integers. For example, the Fibonacci sequence modulo $m = 10$ begins

$$0, 1, 1, 2, 3, 5, 8, 3, 1, 4, 5, 9, 4, 3, 7, 0, 7, 7, 4, 1, 5, 6, 1, 7, 8, 5, \ldots,$$

with the n-th term simply the last digit of F_n. We can also view such a sequence as having terms in $\mathbb{Z}_m = \mathbb{Z}/\langle m \rangle$, the ring of integers modulo m. This has the advantage

MSC2000: 11B50, 11C08, 11T06.

Keywords: linear homogeneous recurrence relations, polynomial congruences, finite rings, finite fields.

that, rather than computing each F_n and dividing that term by m, we can merely begin with 0 and 1 and calculate successive terms of the sequence by adding the two preceding terms in $\mathbb{Z}_m$. This viewpoint makes it obvious that if there is a positive integer ℓ for which $F_\ell = 0$ and $F_{\ell+1} = 1$ (in $\mathbb{Z}_m$), the sequence will then repeat the pattern of $F_0, F_1, \dots, F_{\ell-1}$ indefinitely. For the Fibonacci sequence, it is known that such a value of ℓ exists for every positive integer m. (For $m = 10$, it can be verified that $\ell = 60$.)

Upper limits on the period length of the Fibonacci sequence modulo prime numbers are implicit in Theorem 180 of [Hardy and Wright 1979], one proof of which employs properties of the powers of a root of $f(x) = x^2 - x - 1$, the characteristic polynomial of the Fibonacci sequence. Expanding on this approach, when considering recursive sequences of arbitrary order in this article, we work in rings, $\mathbb{Z}_m[\omega]$, of integers modulo m with a purely formal root ω of the characteristic polynomial f of the sequence adjoined. Our first main result (Corollary 5) is that under minor restrictions on m and the initial terms of the sequence, the period of the recursive sequence modulo m is equal to the order of ω in the group of units in $\mathbb{Z}_m[\omega]$.

Possible orders of ω in the group of units $\mathbb{Z}_p[\omega]^\times$, where p is prime, are determined by the factorization of f in the polynomial ring $\mathbb{Z}_p[x]$. In particular, using properties of ring automorphisms of $\mathbb{Z}_p[\omega]$, we find in Theorem 9 that if f has no repeated factors in $\mathbb{Z}_p[x]$, and t is the least common multiple of the degrees of the irreducible factors of f in $\mathbb{Z}_p[x]$, then t is the smallest positive integer for which the order of ω divides $p^t - 1$. For the Fibonacci sequence, and for other second-order recursive sequences, the important details of the factorization are obtained from standard results about quadratic congruences (particularly calculation of Legendre symbols via the quadratic reciprocity theorem). For sequences of higher order, with characteristic polynomials of higher degree, methods of determining this factorization are less apparent. Finally though, reversing the approach taken with second-order sequences, we show, in Theorem 11 and its corollaries, that information about powers of ω in the rings $\mathbb{Z}_p[\omega]$ lead to highly efficient algorithms for determining the factorization types of monic polynomials f with $\deg f \leq 5$ modulo most primes p.

To outline this article: In Section 2, we define recursive sequences of order k, we consider the simple but instructive case in which $k = 1$, and we establish a criterion for periodicity of recursive sequences modulo arbitrary positive integers m. We introduce the characteristic polynomial f of a recursive sequence in Section 3, which we use to define the rings $\mathbb{Z}_m[\omega]$ referred to above. We show that the periodicity of recursive sequences modulo m can be easily described in terms of powers of the element ω in the ring $\mathbb{Z}_m[\omega]$. This leads us, in Section 4, to consider algebraic properties of these rings. We find that, for a prime modulus p, the relevant properties depend on the factorization of f (e.g., the degrees of irreducible factors, existence of repeated factors) in the ring of polynomials $\mathbb{Z}_p[x]$. In Section 5, we apply well known properties of quadratic congruences to obtain general results about periodicity modulo primes when $k = 2$, with the Fibonacci sequence as a special case. Finally, in Section 6, we obtain efficient algorithms for finding the factorization type of cubic, quartic, and quintic polynomials f modulo most primes p, using calculation of periods of recursive sequences modulo p, or computation of powers of ω. (Adams [1984] and Sun [2003] have separately used certain recursive sequences to develop algorithms for factorization of cubic and quartic polynomials modulo primes. Our algorithm differs in details from both of these.)

The authors are grateful to the referee for pointing out several sources of which we were not aware during the preparation of this article. Engstrom [1931], Ward [1933], and Fillmore and Marx [1968] have extensive details on linear recurrence relations modulo positive integers. See in particular Chapter 8 of [Lidl and Niederreiter 1983] for more results and notes about this aspect of the problem. Furthermore, Skolem [1952] has provided criteria for the factorization type of quartic polynomials modulo primes, similar to our result in Corollary 13, and [Sun 2006] notes a criterion for the factorization of a polynomial into linear factors modulo a prime number, which is essentially the same as the statement of part (1) in our Theorem 15.

2. Periodicity of recursive sequences modulo integers

Let m be a positive integer. We say that a sequence $\{a_n\}_{n=0}^{\infty}$ of integers is *periodic modulo m* or *ℓ-periodic modulo m* if there is a positive integer ℓ such that $a_{\ell+i} \equiv a_i \pmod{m}$ for all $i \geq 0$. We also say that $\{a_n\}_{n=0}^{\infty}$ is periodic in $\mathbb{Z}_m$ in this case, and when it is clear that we are referring to equality in this ring, we write $a_{\ell+i} = a_i$ rather than $a_{\ell+i} \equiv a_i \pmod{m}$. If ℓ is the smallest positive integer for which $\{a_n\}_{n=0}^{\infty}$ is ℓ-periodic modulo m, we call ℓ the *period* of the sequence modulo m.

Proposition 1. *If a sequence $\{a_n\}_{n=0}^{\infty}$ is periodic modulo m with period ℓ, then for a positive integer k, the sequence is k-periodic modulo m if and only if ℓ divides k.*

Proof. Suppose that $\{a_n\}_{n=0}^{\infty}$ is periodic in $\mathbb{Z}_m$ with period ℓ. Then $a_{2\ell+i} = a_{\ell+(\ell+i)} = a_{\ell+i} = a_i$ for all i, and inductively, $a_{\ell q+i} = a_i$ for all positive integers q. So if ℓ divides $k > 0$, then $\{a_n\}_{n=0}^{\infty}$ is k-periodic in $\mathbb{Z}_m$. Conversely then, suppose that $\{a_n\}_{n=0}^{\infty}$ is k-periodic in $\mathbb{Z}_m$ for some positive integer k. We can write $k = \ell q + r$ for some integers q and r with $0 \leq r < \ell$. Now for every $i \geq 0$, we have $a_i = a_{k+i} = a_{\ell q+(r+i)} = a_{r+i}$, since, as noted above, the sequence is ℓq-periodic. If $r > 0$, this contradicts the definition of ℓ as the period of the sequence. So we must conclude that $r = 0$ and so that ℓ divides k. □

In this article, we are primarily interested in the periodicity of sequences defined recursively. We fix the following notation for the sequences of interest. Let k be a positive integer, let $r_1, r_2, \dots, r_k$ be integers, and let $(a_0, a_1, \dots, a_{k-1})$ be a k-tuple of integers. Define a sequence of integers $\{a_n\}_{n=0}^{\infty}$ by setting

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_{k-1} a_{n-k+1} + r_k a_{n-k} = \sum_{i=1}^{k} r_i a_{n-i}, \qquad (2\text{-}1)$$

when $n \geq k$. A sequence of this form is called a *linear homogeneous recurrence relation of order* k; we will refer to it as a *recursive sequence of order* k for short. We call $r_1, r_2, \dots, r_k$ the *coefficients*, and $a_0, a_1, \dots, a_{k-1}$ the *initial terms* of this recursive sequence.

Remark. To establish that $\{a_n\}_{n=0}^{\infty}$ as defined in (2-1) is ℓ-periodic in $\mathbb{Z}_m$, it suffices, as we noted in Section 1 for the Fibonacci sequence, to show that $a_{\ell+i} = a_i$ for $0 \leq i \leq k-1$.

We can describe the periodicity of recursive sequences of order $k = 1$ using standard results about linear congruences from elementary number theory.

Example. Define a_n for $n \geq 0$ by setting $a_n = r a_{n-1}$ when $n > 0$, with r and a_0 integers. Then $a_n = a_0 r^n$ for all n, and the sequence is periodic modulo m if there is a positive integer ℓ such that $a_0 r^{\ell} \equiv a_0 \pmod{m}$. If $\gcd(a_0, m) = d$, this congruence is equivalent to $r^{\ell} \equiv 1 \pmod{m/d}$, and such a value of ℓ exists if and only if r is relatively prime to m/d. In that case, the period of the sequence equals $\text{ord}_{m/d}(r)$, the order of r in the group $\mathbb{Z}_{m/d}^{\times}$ of units in $\mathbb{Z}_{m/d}$.

Remark. This example illustrates that we are unlikely to obtain a precise formula for the period of a recursive sequence modulo every positive integer m. For example, if $a_0 = 1$ and $a_n = 2a_{n-1}$ for $n > 0$, then the sequence $\{a_n\}_{n=0}^{\infty}$ is periodic modulo every odd positive integer m, with period the order of 2 in $\mathbb{Z}_m^{\times}$. We know that this order divides $\phi(m) = |\mathbb{Z}_m^{\times}|$, but a more specific formula for this value is difficult to obtain. Similarly, for larger values of k, we will generally be able to provide only upper limits on the period of a recursive sequence modulo an arbitrary integer m.

The following theorem provides a criterion for the periodicity of recursive sequences modulo positive integers m. Our proof follows that of a similar result in [Wall 1960] for the Fibonacci sequence.

Theorem 2. *Let $\{a_n\}_{n=0}^{\infty}$ be a recursive sequence with coefficients $r_1, r_2, \dots, r_k$, defined as in (2-1). Let m be a positive integer. If* $\gcd(r_k, m) = 1$, *then the sequence is periodic modulo m.*

Proof. There are m^k distinct k-tuples of elements of $\mathbb{Z}_m$. By the pigeonhole principle, it follows that there are integers s and t with $0 \leq s < t \leq m^k$ such that $a_{s+i} = a_{t+i}$

in $\mathbb{Z}_m$ for $0 \le i \le k-1$. We may assume that s is the smallest nonnegative integer for which this is true. But if $s > 0$, then $a_{s+k-1} = a_{t+k-1}$ implies that

$$\begin{aligned} r_1 a_{s+k-2} + r_2 a_{s+k-3} + \cdots + r_{k-1} a_s + r_k a_{s-1} \\ = r_1 a_{t+k-2} + r_2 a_{t+k-3} + \cdots + r_{k-1} a_t + r_k a_{t-1} \end{aligned}$$

in $\mathbb{Z}_m$, by the recursive definition of the sequence. It follows that $r_k a_{s-1} = r_k a_{t-1}$, and if $\gcd(r_k, m) = 1$, so that r_k is a unit in $\mathbb{Z}_m$, then $a_{s-1} = a_{t-1}$ in $\mathbb{Z}_m$. This contradicts our assumption about s, so we must conclude that $s = 0$. By the note above, it follows that $\{a_n\}_{n=0}^{\infty}$ is periodic modulo m. □

Remark. If $\gcd(r_k, m) > 1$, then $\{a_n\}_{n=0}^{\infty}$ defined by (2-1) may or may not be periodic modulo m, depending on the initial terms of the sequence. For example, if $(a_0, a_1, \ldots, a_{k-1}) = (1, 0, \ldots, 0)$, then it is easy to see that r_k divides a_n for all $n > 0$, and so $a_\ell \equiv a_0 \pmod{m}$ is not possible for any $\ell > 0$. On the other hand, the sequence with initial terms $(a_0, a_1, \ldots, a_{k-1}) = (0, 0, \ldots, 0)$ is clearly 1-periodic modulo m. This trivial example is generally not exclusive. For instance, if $a_n = a_{n-1} + a_{n-2} + 2a_{n-3}$, with $(a_0, a_1, a_2) = (1, 0, 1)$, then the sequence $\{a_n\}_{n=0}^{\infty}$ is 3-periodic modulo $m = 2$. In any event, the proof of Theorem 2 shows that every recursive sequence defined as in (2-1) will exhibit an infinitely repeating pattern of terms modulo m, possibly following some initial terms. In the remainder of this article, given a recursive sequence of order k, we will restrict our attention to moduli m that are relatively prime to the k-th order coefficient r_k.

3. Polynomial extensions of $\mathbb{Z}_m$

If $\{a_n\}_{n=0}^{\infty}$ is a recursive sequence given as in (2-1), then we define the *characteristic polynomial* of that sequence to be

$$f(x) = x^k - r_1 x^{k-1} - r_2 x^{k-2} - \cdots - r_{k-1} x - r_k.$$

It is well known that each a_n can be expressed in terms of n-th powers of the solutions of $f(x) = 0$, with the combination of those powers determined by the initial terms of the sequence. In considering arithmetic properites of the sequence $\{a_n\}_{n=0}^{\infty}$ modulo m, we will find it useful to work in rings, $\mathbb{Z}_m[\omega]$, of the integers modulo m with a purely formal solution, ω, of $f(x) = 0$ adjoined. We define these rings as follows.

For a positive integer m, consider the quotient ring $\mathbb{Z}_m[x]/\langle f \rangle$, where $\mathbb{Z}_m[x]$ is the ring of polynomials with coefficients in $\mathbb{Z}_m$ and $\langle f \rangle$ is the principal ideal of $\mathbb{Z}_m[x]$ generated by f. Since f is a *monic* polynomial, that is, its leading coefficient is 1, then for every polynomial g in $\mathbb{Z}_m[x]$, there exist unique polynomials q and r in $\mathbb{Z}_m[x]$ such that $g = f \cdot q + r$, with r of smaller degree than f, or $r = 0$. In that case, $g + \langle f \rangle = r + \langle f \rangle$. Writing the coset $x + \langle f \rangle$ as ω_f, or as ω when f is apparent from context, we can identify $\mathbb{Z}_m[x]/\langle f \rangle$ with the ring $\mathbb{Z}_m[\omega]$ defined by

$$\mathbb{Z}_m[\omega] = \Big\{ b_{k-1}\omega^{k-1} + b_{k-2}\omega^{k-2} + \cdots + b_1\omega + b_0 \,\Big|\, b_i \in \mathbb{Z}_m \text{ and } \omega^k = \sum_{i=1}^{k} r_i \omega^{k-i} \Big\}. \tag{3-1}$$

Here $b_{k-1}\omega^{k-1} + \cdots + b_0 = c_{k-1}\omega^{k-1} + \cdots + c_0$ if and only if $b_i = c_i$ in $\mathbb{Z}_m$ for $0 \le i \le k-1$, so in general, $\mathbb{Z}_m[\omega]$ has m^k elements. We refer to $\mathbb{Z}_m[\omega]$ as the *extension of $\mathbb{Z}_m$ by the polynomial f* or more generally as a *polynomial extension* of $\mathbb{Z}_m$. We write elements of $\mathbb{Z}_m[\omega]$ using Greek letters, or in the form $g(\omega)$ where g is a polynomial in $\mathbb{Z}_m[x]$.

We establish a connection between the ring $\mathbb{Z}_m[x]/\langle f \rangle$ and recursive sequences with characteristic polynomial f as follows. Let $\{a_n\}_{n=0}^{\infty}$ be defined as in (2-1), and for $1 \le j \le k$ and $n \ge k$, let $a(j, n) = \sum_{i=j}^{k} r_i a_{n-i}$. Notice that, for all $n \ge k$,

$$a(k, n) = r_k a_{n-k} \tag{3-2}$$

and

$$a(j+1, n) + r_j a_{n-j} = a(j, n) \quad \text{if } 1 \le j < k. \tag{3-3}$$

Now define α to be the following element of $\mathbb{Z}_m[\omega]$, determined by the initial terms and coefficients of the sequence:

$$\begin{aligned}\alpha &= a_{k-1}\omega^{k-1} + a(2,k)\omega^{k-2} + a(3,k+1)\omega^{k-3} + \cdots + a(k-1,2k-3)\omega + a(k,2k-2)\\ &= a_{k-1}\omega^{k-1} + \sum_{j=2}^{k} a(j, k+j-2) \cdot \omega^{k-j},\end{aligned} \tag{3-4}$$

here viewing a_{k-1} and each $a(j, k+j-2)$ as elements of $\mathbb{Z}_m$.

Theorem 3. *Let $\{a_n\}_{n=0}^{\infty}$ be defined recursively as in (2-1), and let α be defined by (3-4). Then for every integer $n \ge 0$,*

$$\alpha\omega^n = a_{n+k-1}\omega^{k-1} + \sum_{j=2}^{k} a(j, n+k+j-2) \cdot \omega^{k-j}. \tag{3-5}$$

Remark. If $n \ge 1$, then $a_{n+k-1} = \sum_{i=1}^{k} r_i a_{n+k-1-i} = a(1, n+k-1)$ by the recursive definition of the sequence. So for $n \ge 1$, we can also express (3-5) as

$$\alpha\omega^n = \sum_{j=1}^{k} a(j, n+k+j-2) \cdot \omega^{k-j}. \tag{3-6}$$

Proof. We use induction on n. Equation (3-5) is true for $n = 0$ by (3-4). So suppose that (3-5) holds for some integer $n \ge 0$. Then

$$\begin{aligned}\alpha\omega^{n+1} = (\alpha\omega^n)\omega &= a_{n+k-1}\omega^k + \sum_{j=2}^{k} a(j, n+k+j-2)\cdot\omega^{k-j+1}\\ &= \sum_{j=1}^{k} r_j a_{n+k-1}\cdot\omega^{k-j} + \sum_{j=2}^{k} a(j, n+k+j-2)\cdot\omega^{k-j+1},\end{aligned}$$

using the equation for ω^k in (3-1). Splitting off the last term in the first sum, and replacing j by $j+1$ in the second sum, we have that

$$\begin{aligned}\alpha\omega^{n+1} &= r_k a_{n+k-1} + \sum_{j=1}^{k-1} r_j a_{n+k-1}\cdot\omega^{k-j} + \sum_{j=1}^{k-1} a(j+1, n+k+j-1)\cdot\omega^{k-j}\\ &= r_k a_{n+k-1} + \sum_{j=1}^{k-1}(r_j a_{n+k-1} + a(j+1, n+k+j-1))\cdot\omega^{k-j}\\ &= r_k a_{n+k-1} + \sum_{j=1}^{k-1} a(j, n+k+j-1)\cdot\omega^{k-j},\end{aligned}$$

using (3-3). But $r_k a_{n+k-1} = a(k, n+2k-1)$ by (3-2), so that

$$\alpha\omega^{n+1} = \sum_{j=1}^{k} a(j, n+k+j-1)\cdot\omega^{k-j}.$$

This is (3-6) with $n+1$ in place of n. Since $n+1 \geq 1$, (3-5) is then true with $n+1$ in place of n, and so (3-5) holds for all integers $n \geq 0$ by induction. □

Theorem 4. *Let k be a positive integer, and let $\{a_n\}_{n=0}^{\infty}$ be a recursive sequence with coefficients $r_1, \dots, r_k$ and characteristic polynomial f, defined as in* (2-1)*. Let m be a positive integer such that $\gcd(r_k, m) = 1$, let $\mathbb{Z}_m[\omega] = \mathbb{Z}_m[x]/\langle f\rangle$, and let α be given as in* (3-4)*. Then $\{a_n\}_{n=0}^{\infty}$ is ℓ-periodic modulo m if and only if $\alpha\omega^\ell = \alpha$ in $\mathbb{Z}_m[\omega]$.*

Proof. If $a_{\ell+i} = a_i$ for all $i \geq 0$, then, in particular, $a_{\ell+k-1} = a_{k-1}$, and it is easy to see that $a(j, \ell+k+j-2) = a(j, k+j-2)$ for $2 \leq j \leq k$. Thus $\alpha\omega^\ell = \alpha$ by (3-5).

Conversely, suppose that $\alpha\omega^\ell = \alpha$. Comparing the equations in (3-4) and (3-5), we know that $a_{\ell+k-1} = a_{k-1}$ and $a(j, \ell+k+j-2) = a(j, k+j-2)$ in $\mathbb{Z}_m$ for $2 \leq j \leq k$. But if $\gcd(r_k, m) = 1$, so that r_k is a unit in $\mathbb{Z}_m$, we can use the latter equations to show inductively that $a_{\ell+j-2} = a_{j-2}$ for $2 \leq j \leq k$, which is sufficient to establish that the sequence is ℓ-periodic. If $j = k$, then $a(k, \ell+2k-2) = a(k, 2k-2)$ implies that $r_k a_{\ell+k-2} = r_k a_{k-2}$, so that $a_{\ell+k-2} = a_{k-2}$. Now let j be an integer with $2 \leq j < k$, and suppose that we have shown that $a_{\ell+i-2} = a_{i-2}$ for $j < i \leq k$.

Then $a(j, \ell+k+j-2) = a(j, k+j-2)$ implies that

$$\begin{aligned} &r_j a_{\ell+k-2} + r_{j+1} a_{\ell+k-3} + \cdots + r_{k-1} a_{\ell+j-1} + r_k a_{\ell+j-2} \\ &\qquad\qquad\qquad = r_j a_{k-2} + r_{j+1} a_{k-3} + \cdots + r_{k-1} a_{j-1} + r_k a_{j-2}, \end{aligned}$$

which, by the inductive hypothesis and the assumption that r_k is a unit, implies that $a_{\ell+j-2} = a_{j-2}$. The result follows by induction. □

Corollary 5. *Let k be a positive integer, and let $\{a_n\}_{n=0}^{\infty}$ be a recursive sequence with coefficients $r_1, \dots, r_k$ and characteristic polynomial f, defined as in (2-1). Let m be a positive integer such that $\gcd(r_k, m) = 1$, let $\mathbb{Z}_m[\omega] = \mathbb{Z}_m[x]/\langle f\rangle$, and let α be given as in (3-4). Then ω is a unit in $\mathbb{Z}_m[\omega]$, and $\{a_n\}_{n=0}^{\infty}$ is periodic modulo m, with period ℓ dividing $\operatorname{ord}_m(\omega)$, the order of ω in the group, $\mathbb{Z}_m[\omega]^{\times}$, of units in $\mathbb{Z}_m[\omega]$. If $A = \{\beta \in \mathbb{Z}_m[\omega] \mid \alpha\beta = 0\}$, then ℓ is the order of $\omega + A$ in the group of units of the quotient ring $\mathbb{Z}_m[\omega]/A$.*

Remark. It is easy to see that the set A defined in the corollary is an ideal of $\mathbb{Z}_m[\omega]$. This ideal, called the *annihilator* of α in $\mathbb{Z}_m[\omega]$, is trivial if α is a unit in $\mathbb{Z}_m[\omega]$, so in that case, $\ell = \operatorname{ord}_m(\omega)$.

Proof. If $\gcd(r_k, m) = 1$, then r_k is a unit in $\mathbb{Z}_m$, say with inverse r_k^{-1}. Then it is easy to verify that $r_k^{-1}(\omega^{k-1} - r_1\omega^{k-2} - \cdots - r_{k-2}\omega - r_{k-1}) \cdot \omega = 1$, so that ω is a unit in $\mathbb{Z}_m[\omega]$. Since $\mathbb{Z}_m[\omega]^{\times}$ is finite, there is an integer $t = \operatorname{ord}_m(\omega)$ for which $\omega^t = 1$. But then $\alpha\omega^t = \alpha$, and Theorem 4 implies that $\{a_n\}_{n=0}^{\infty}$ is t-periodic modulo m. If ℓ is the period of this sequence modulo m, we know that ℓ divides t by Proposition 1. Furthermore, ℓ is the smallest positive integer such that $\alpha\omega^{\ell} = \alpha$, which is true if and only if $\omega^{\ell} - 1$ is in the annihilator of α. But then ℓ is the order of $\omega + A$ as a unit in the quotient ring $\mathbb{Z}_m[\omega]/A$. □

Example. Consider the recursive sequence of order $k = 1$ defined by $a_n = ra_{n-1}$ for $n > 0$, with a_0 and r fixed integers, as in a previous example. Let m be a positive integer that is relatively prime to r, in which case the sequence is periodic modulo m. The characteristic polynomial of $\{a_n\}_{n=0}^{\infty}$ is $f(x) = x - r$, so that $\omega = x + \langle f\rangle = r + \langle f\rangle$ in $\mathbb{Z}_m[x]/\langle f\rangle$. It is easy to see that $\mathbb{Z}_m[x]/\langle f\rangle$ is isomorphic to $\mathbb{Z}_m$, so that we can identify ω with r. By (3-4), we have that $\alpha = a_0$, and if $\gcd(a_0, m) = d$, then we find that the annihilator A of α in $\mathbb{Z}_m[\omega]$ is generated by m/d. Corollary 5 implies that the period of $\{a_n\}_{n=0}^{\infty}$ is the order of $r + A$ in $(\mathbb{Z}_m[\omega]/A)^{\times}$, which we can view as the order of r in $\mathbb{Z}_{m/d}^{\times}$. Thus we see that Corollary 5 generalizes our results for recursive sequences of order $k = 1$ to higher orders.

Example. It can be verified that the period of the Fibonacci sequence modulo $m = 5$ is 20. On the other hand, the *Lucas sequence*, defined for $n \geq 0$ by $(L_0, L_1) = (2, 1)$, and $L_n = L_{n-1} + L_{n-2}$ if $n > 1$, has period four modulo $m = 5$. This is possible because, for the Fibonacci sequence, $\alpha = \omega$ is a unit in $\mathbb{Z}_5[\omega]$, where $\omega^2 = \omega + 1$, while for the Lucas sequence, $\alpha = \omega + 2$ has a nontrivial annihilator in $\mathbb{Z}_5[\omega]$.

Remark. If the initial terms of a recursive sequence are $(a_0, \dots, a_{k-2}, a_{k-1}) = (0, \dots, 0, 1)$, then $\alpha = \omega^{k-1}$ is a unit, when ω is a unit in $\mathbb{Z}_m[\omega]$. In this case, Corollary 5 implies that the period of the sequence modulo m is the same as the order of ω in $\mathbb{Z}_m[\omega]^\times$. We will restrict our attention to this special case for the initial terms in what follows.

In the remainder of this article, we will further restrict our attention to the case in which the modulus m of interest is prime, using the following observations. First suppose that $\{a_n\}_{n=0}^\infty$ is periodic modulo s with period k, and periodic modulo t with period ℓ. If $\gcd(s, t) = 1$, it is straightforward to show, using Proposition 1, that $\{a_n\}_{n=0}^\infty$ is periodic modulo st with period $\operatorname{lcm}(k, \ell)$. (This does not require the assumption that the sequence is defined recursively.) For powers of primes, we can invoke the following result.

Theorem 6. *Let p be a prime number and j a positive integer. Let f be a polynomial with integer coefficients, and suppose that p does not divide the constant coefficient of f, so that ω is a unit in $\mathbb{Z}_{p^j}[\omega] = \mathbb{Z}_{p^j}[x]/\langle f\rangle$. Let $s = \operatorname{ord}_{p^j}(\omega)$ and $t = \operatorname{ord}_{p^{j+1}}(\omega)$. Then either $t = s$ or $t = ps$.*

Remark. If d divides m, then it is easy to see that the function $\phi : \mathbb{Z}_m[\omega] \to \mathbb{Z}_d[\omega]$ defined by $\phi(g(\omega)) = g(\omega)$ is a well-defined ring homomorphism, with kernel $\langle d\rangle$. So if $g(\omega) = h(\omega)$ in $\mathbb{Z}_m[\omega]$, then $g(\omega) = h(\omega)$ in $\mathbb{Z}_d[\omega]$. On the other hand, if $g(\omega) = h(\omega)$ in $\mathbb{Z}_d[\omega]$, then the strongest statement that we can make is that $g(\omega) = h(\omega) + d \cdot \delta$ for some element δ in $\mathbb{Z}_m[\omega]$.

Proof. Let s be the order of ω in $\mathbb{Z}_{p^j}[\omega]$ and let t be the order of ω in $\mathbb{Z}_{p^{j+1}}[\omega]$. Since $\omega^t = 1$ in $\mathbb{Z}_{p^{j+1}}[\omega]$, then $\omega^t = 1$ in $\mathbb{Z}_{p^j}[\omega]$ by the remark above, so that s divides t. By the same remark, since $\omega^s = 1$ in $\mathbb{Z}_{p^j}[\omega]$, then $\omega^s = 1 + p^j \cdot \delta$ for some δ in $\mathbb{Z}_{p^{j+1}}[\omega]$. But now

$$\omega^{ps} = (\omega^s)^p = (1 + p^j \cdot \delta)^p = 1 + \tbinom{p}{1} p^j \cdot \delta + \tbinom{p}{2} p^{2j} \cdot \delta^2 + \cdots + p^{pj} \cdot \delta^p = 1$$

in $\mathbb{Z}_{p^{j+1}}[\omega]$, since all terms in the sum aside from the first are divisible by p^{j+1}. Thus t divides ps. Since $s \mid t$ and $t \mid ps$, with p prime, we conclude that $t = s$ or $t = ps$. □

So if ℓ is the period of a recursive sequence modulo p, then the period of the same sequence modulo p^j must divide $p^{j-1} \cdot \ell$. Interesting questions about periods of recursive sequences modulo prime powers remain open. For example, Sun and Sun [1992] showed that if a prime exponent p were a counterexample to the first case of Fermat's Last Theorem, then the period of the Fibonacci sequence modulo

p and modulo p^2 would have to be the same. It is not known whether any such primes exist for the Fibonacci sequence. (Of course, it is now known that no such counterexamples to Fermat's Last Theorem can exist.) For our purposes, we will simply note that the upper limit given above is not always obtained as the exact period of a recursive sequence modulo p^j, as the following example shows.

Example. Define a_n for $n \geq 0$ by $(a_0, a_1, a_2) = (0, 0, 1)$ and

$$a_n = a_{n-1} + a_{n-2} + 2a_{n-3}$$

for $n > 2$. We find that $\{a_n\}_{n=0}^{\infty}$ has period $\ell = 6$ both modulo $p = 3$ and modulo $p^2 = 9$.

4. Algebraic properties of $\mathbb{Z}_p[\omega]$

With these restrictions in place, our main task, given the characteristic polynomial f of a recursive sequence, is to describe the order of $\omega = \omega_f = x + \langle f \rangle$ as a unit in the quotient ring $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$, for all primes p not dividing the constant coefficient of f. We will see that our description of $\text{ord}_p(\omega)$ depends largely on how f factors in the polynomial ring $\mathbb{Z}_p[x]$. We begin by compiling some useful general statements about these polynomial extensions.

(1) If g divides f, then the function $\phi : \mathbb{Z}_p[\omega_f] \to \mathbb{Z}_p[\omega_g]$ defined by $\phi(h(\omega_f)) = h(\omega_g)$ is a well-defined ring homomorphism with kernel $\langle g(\omega_f)\rangle$. It follows that if $r(\omega_f) = s(\omega_f)$ in $\mathbb{Z}_p[\omega_f]$, then $r(\omega_g) = s(\omega_g)$ in $\mathbb{Z}_p[\omega_g]$, while if $r(\omega_g) = s(\omega_g)$ in $\mathbb{Z}_p[\omega_g]$, then $r(\omega_f) = s(\omega_f) + g(\omega_f) \cdot \delta$ for some δ in $\mathbb{Z}_p[\omega_f]$.

(2) The set of all (ring) automorphisms of $\mathbb{Z}_p[\omega]$ forms a group under composition. If h is a polynomial in $\mathbb{Z}_p[x]$ and $\sigma : \mathbb{Z}_p[\omega] \to \mathbb{Z}_p[\omega]$ is an automorphism, then $\sigma(h(\omega)) = h(\sigma(\omega))$. In particular, $0 = \sigma(0) = \sigma(f(\omega)) = f(\sigma(\omega))$, so that $\sigma(\omega)$ is a root of f.

(3) For an automorphism σ of $\mathbb{Z}_p[\omega]$, if $\sigma(\omega) = \omega$, then $\sigma(h(\omega)) = h(\sigma(\omega)) = h(\omega)$ for all $h \in \mathbb{Z}_p[x]$. That is, $\sigma(\omega) = \omega$ if and only if σ is the identity automorphism.

(4) The function $\sigma_p : \mathbb{Z}_p[\omega] \to \mathbb{Z}_p[\omega]$ defined by $\sigma_p(\beta) = \beta^p$ is a ring homomorhism, since $\mathbb{Z}_p[\omega]$ has characteristic p. Furthermore, σ_p is an automorphism if and only if the polynomial f has no repeated irreducible factors in $\mathbb{Z}_p[x]$. (If $f = g^2 h$ for some irreducible polynomial g, then $g(\omega)h(\omega)$ is a nonzero element in the kernel of σ_p. On the other hand, if f has no repeated irreducible factors, then the uniqueness of irreducible factorization in $\mathbb{Z}_p[x]$ shows that f divides h^p if and only if f divides h. In that case, the kernel of σ_p is trivial, and since $\mathbb{Z}_p[\omega]$ is finite, σ_p is a bijection.)

(5) If f is irreducible in $\mathbb{Z}_p[x]$, with $\deg f = k$, then $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$ is a field with p^k elements. In this case, the group $\mathrm{Aut}(\mathbb{Z}_p[\omega])$ of automorphisms of $\mathbb{Z}_p[\omega]$ is cyclic of order k, generated by σ_p [Dummit and Foote 2004, p. 556].

(6) If $f = f_1 \cdot f_2 \cdots f_j$ is a product of pairwise relatively prime polynomials in $\mathbb{Z}_p[x]$, then the quotient ring $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$ is isomorphic to the direct product of quotient rings $\mathbb{Z}_p[x]/\langle f_1 \rangle \times \mathbb{Z}_p[x]/\langle f_2 \rangle \times \cdots \times \mathbb{Z}_p[x]/\langle f_j \rangle$ [Dummit and Foote 2004, p. 313].

We can draw some conclusions about the order of ω in $\mathbb{Z}_p[\omega]^\times$ from these statements. We begin with the case in which f is irreducible in $\mathbb{Z}_p[x]$.

Theorem 7. *Let $f(x) = x^k - r_1 x^{k-1} - \cdots - r_k$. Let p be a prime for which $p \nmid r_k$, and suppose that f is irreducible in $\mathbb{Z}_p[x]$. Let t be the order of $(-1)^{k+1} r_k$ as an element of $\mathbb{Z}_p^\times$. Then $\mathrm{ord}_p(\omega)$, the order of ω as a unit in $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$, divides $\frac{p^k-1}{p-1} t$, but $\mathrm{ord}_p(\omega)$ divides neither $p^i - 1$ for $0 < i < k$ nor $\frac{p^k-1}{p-1} s$ for $0 < s < t$.*

Proof. By statement (5), we know that $\mathrm{Aut}(\mathbb{Z}_p[\omega])$ is cyclic of order k, generated by σ_p. The composition of i copies of σ_p is the same as σ_{p^i}, defined by $\sigma_{p^i}(\beta) = \beta^{p^i}$. Statement (3) implies that $\omega^{p^i} \neq \omega$, and so $\omega^{p^i - 1} \neq 1$, for $0 < i < k$.

Statement (2) now implies that f has k distinct roots in $\mathbb{Z}_p[\omega]$, each of the form $\sigma_{p^i}(\omega) = \omega^{p^i}$ for $0 \le i < k$, and therefore

$$f(x) = (x - \omega)(x - \omega^p)(x - \omega^{p^2}) \cdots (x - \omega^{p^{k-1}}).$$

Comparing constant coefficients of these polynomials, we find that $-r_k = (-1)^k \omega \cdot \omega^p \cdot \omega^{p^2} \cdots \omega^{p^{k-1}}$, and so

$$(-1)^{k+1} r_k = \omega^{1+p+p^2+\cdots+p^{k-1}} = \omega^{\frac{p^k-1}{p-1}}.$$

If t is the order of $(-1)^{k+1} r_k$ in $\mathbb{Z}_p^\times$, then $\omega^{\frac{p^k-1}{p-1} t} = 1$, but $\omega^{\frac{p^k-1}{p-1} s} \neq 1$ for $0 < s < t$. □

Secondly, we consider the case in which f is a power of an irreducible polynomial.

Theorem 8. *Let f be a monic polynomial of degree k with integer coefficients. Suppose that $f = g^t$, where g is an irreducible polynomial of degree s in $\mathbb{Z}_p[x]$ (so that $st = k$). Let p be a prime number not dividing the constant coefficient of g (and so not dividing the constant coefficient of f). Let j be the smallest nonnegative integer for which $p^j \ge t$. Let $\mathbb{Z}_p[\omega_f] = \mathbb{Z}_p[x]/\langle f \rangle$ and $\mathbb{Z}_p[\omega_g] = \mathbb{Z}_p[x]/\langle g \rangle$, and suppose that ω_g has order ℓ as a unit in $\mathbb{Z}_p[\omega_g]$. Then the order of ω_f as a unit in $\mathbb{Z}_p[\omega_f]$ equals $p^i \ell$ for some i with $0 \le i \le j$.*

Proof. Let m be the order of ω_f in the group $\mathbb{Z}_p[\omega_f]^\times$. Since $(\omega_f)^m = 1$ in $\mathbb{Z}_p[\omega_f]$, then $(\omega_g)^m = 1$ in $\mathbb{Z}_p[\omega_g]$ by statement (1), so that ℓ divides m. Since $(\omega_g)^\ell = 1$ in $\mathbb{Z}_p[\omega_g]$, statement (1) also implies that $(\omega_f)^\ell = 1 + g(\omega_f)\cdot\delta$ for some δ in $\mathbb{Z}_p[\omega_f]$. Now note that

$$(\omega_f)^{p^j\ell} = ((\omega_f)^\ell)^{p^j} = (1 + g(\omega_f)\cdot\delta)^{p^j} = 1 + g(\omega_f)^{p^j}\cdot\delta^{p^j} = 1,$$

in $\mathbb{Z}_p[\omega_f]$, using the facts that $\mathbb{Z}_p[\omega_f]$ has characteristic p and that $f = g^t$ divides g^{p^j}, by the definition of j. So m divides $p^j\ell$, and the conclusion of Theorem 8 follows immediately. □

Finally, if f factors as a product of pairwise relatively prime polynomials, say $f = f_1 \cdot f_2 \cdots f_j$ with each f_i a power of a distinct irreducible polynomial in $\mathbb{Z}_p[x]$, then $\mathbb{Z}_p[\omega_f]$ is isomorphic to

$$\mathbb{Z}_p[\omega_{f_1}] \times \mathbb{Z}_p[\omega_{f_2}] \times \cdots \times \mathbb{Z}_p[\omega_{f_j}]$$

by statement (6). If a prime number p does not divide the constant coefficient of f, then it is easy to see that the order of ω_f in $\mathbb{Z}_p[\omega_f]^\times$ is the least common multiple of the orders of each ω_{f_i} in the appropriate group of units. We can place a further restriction on the order of ω_f when no irreducible factor of f is repeated.

Theorem 9. *Let f be a monic polynomial of degree k with integer coefficients, and let p be a prime number not dividing the constant coefficient of f. Suppose that $f = f_1 \cdot f_2 \cdots f_j$ for distinct irreducible polynomials f_i of degree k_i in $\mathbb{Z}_p[x]$ (so that $k = k_1 + k_2 + \cdots + k_j$). Let $t = \operatorname{lcm}(k_1, k_2, \ldots, k_j)$. Then in the group $\mathbb{Z}_p[\omega]^\times$ of units in the ring $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$, the order of ω divides $p^t - 1$, but does not divide $p^i - 1$ for $0 < i < t$.*

Proof. By statement (4), the function $\sigma_p : \mathbb{Z}_p[\omega] \to \mathbb{Z}_p[\omega]$ defined by $\sigma_p(\beta) = \beta^p$ is an automorphism of $\mathbb{Z}_p[\omega]$. With $\mathbb{Z}_p[\omega_f]$ isomorphic to $\mathbb{Z}_p[\omega_{f_1}] \times \mathbb{Z}_p[\omega_{f_2}] \times \cdots \times \mathbb{Z}_p[\omega_{f_j}]$ and each $\mathbb{Z}_p[\omega_{f_i}]$ a field, it is straightforward to show that the order of σ_p in $\operatorname{Aut}(\mathbb{Z}_p[\omega])$ is $t = \operatorname{lcm}(k_1, k_2, \ldots, k_j)$. By statement (3), it follows that $\omega^{p^t} = \omega$, but $\omega^{p^i} \neq \omega$ if $0 < i < t$. Since ω is a unit in $\mathbb{Z}_p[\omega]$, the conclusion of Theorem 9 follows. □

5. Recursive sequences of order two

We illustrate our results so far with some general statements about recursive sequences of order two. Define a_n for $n \geq 0$ by $(a_0, a_1) = (0, 1)$, and $a_n = r_1 a_{n-1} + r_2 a_{n-2}$ for $n > 1$, where r_1 and r_2 are integers. Let p be a prime number, let $f(x) = x^2 - r_1 x - r_2$, and let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$. If $p \nmid r_2$, then $\{a_n\}_{n=0}^\infty$ is periodic modulo p, with period ℓ equal to the order of ω in $\mathbb{Z}_p[\omega]^\times$. The factorization of f in $\mathbb{Z}_p[x]$ is determined by its *discriminant*, $D = D(f) = r_1^2 + 4r_2$, and we can use that factorization to describe ℓ.

Case 1: *f is irreducible in $\mathbb{Z}_p[x]$.* For odd p, this is the case if and only if the Legendre symbol $\left(\frac{D}{p}\right)$ equals -1, while for $p = 2$ this occurs precisely when $D \equiv 5 \pmod 8$. Theorem 7 implies that ℓ divides $(p+1)t$, where t is the order of $-r_2$ in $\mathbb{Z}_p^\times$, but that ℓ divides neither $p-1$ nor $(p+1)s$ for $0 < s < t$.

Case 2: *f factors as a product of distinct linear polynomials in $\mathbb{Z}_p[x]$.* For odd p, this is the case if and only if $\left(\frac{D}{p}\right) = 1$, while for $p = 2$, this occurs in general when $D \equiv 1 \pmod 8$. (Of course, it is impossible for a quadratic polynomial f to factor into distinct linear terms in $\mathbb{Z}_2[x]$ unless 2 divdes its constant coefficient, which we assume is not the case here.) Theorem 9 implies that ℓ divides $p-1$. More precisely, if $f(x) = (x-b)(x-c)$ in $\mathbb{Z}_p[x]$, then ℓ is the least common multiple of the orders of b and c in $\mathbb{Z}_p^\times$. (If $f_1(x) = x - b$, then $\omega_{f_1} = x + \langle f_1 \rangle = b + \langle f_1 \rangle$ in $\mathbb{Z}_p[x]/\langle f_1 \rangle$, which is isomorphic to $\mathbb{Z}_p$.)

Case 3: *f factors as the square of a linear polynomial in $\mathbb{Z}_p[x]$.* This is the case if and only if p divides D. Since $p \geq 2$ for every prime p, Theorem 8 implies that ℓ divides $p(p-1)$. In this case, we can make the following precise statement as a corollary of Theorem 8.

Corollary 10. *Let $f(x) = x^2 - r_1 x - r_2$ with r_1 and r_2 integers. Let p be a prime number dividing $D = r_1^2 + 4r_2$ but not dividing r_2, so that $f(x) = (x-c)^2$ in $\mathbb{Z}_p[x]$ for some $c \neq 0$ in $\mathbb{Z}_p$. If t is the order of c in $\mathbb{Z}_p^\times$, then the order of $\omega = x + \langle f \rangle$ as a unit in $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$ is pt.*

Proof. Let $g(x) = x - c$, and let t be the order of c in $\mathbb{Z}_p^\times$. Since $\omega_g = x + \langle g \rangle = c + \langle g \rangle$, then t is the order of ω_g as a unit in $\mathbb{Z}_p[\omega_g]$. Theorem 8 implies that the order of ω_f in $\mathbb{Z}_p[\omega_f]^\times$ is either t or pt. But $(\omega_f)^t = 1$ if and only if $f(x) = (x-c)^2$ divides $h(x) = x^t - 1$ in $\mathbb{Z}_p[x]$. If so, then $h(c)$ and $h'(c)$ are both zero in $\mathbb{Z}_p$. This is impossible since $h'(c) = tc^{t-1}$, but $p \nmid t$ (a divisor of $p-1$) and $p \nmid c$. So the order of ω_f in $\mathbb{Z}_p[\omega_f]^\times$ must be pt. □

Example. For the Fibonacci sequence, $r_1 = 1$, $r_2 = 1$, and $D = 5$. Since $p \nmid r_2$ for all primes p, the Fibonacci sequence is periodic modulo p, say with period ℓ_p. The polynomial $x^2 - x - 1$ is irreducible in $\mathbb{Z}_2[x]$, since $D \equiv 5 \pmod 8$. The order of $-r_2 = -1$ in $\mathbb{Z}_2^\times$ is 1, and so for $p = 2$, we have that ℓ_p is a divisor of $p + 1 = 3$, but not $p - 1 = 1$. The only possibility is $\ell_2 = 3$, which is easy to verify directly. Since $x^2 - x - 1 = (x-3)^2$ in $\mathbb{Z}_5[x]$, and $c = 3$ has order four in $\mathbb{Z}_5^\times$, Corollary 10 implies that $\ell_5 = 20$. (Note that these two results, together with the remark preceding Theorem 6, verify the claim made in the introduction that the Fibonacci sequence has period $\ell = 60$ modulo $m = 10$.)

If $p \neq 2, 5$, then since $5 \equiv 1 \pmod 4$, quadratic reciprocity implies that $\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right)$, so that factorization of $x^2 - x - 1$ is determined by the value of p modulo 5. If $p \equiv 1$ or $4 \pmod 5$, then $\left(\frac{5}{p}\right) = 1$ and $x^2 - x - 1$ factors as a product of linear factors in $\mathbb{Z}_p[x]$. Theorem 9 implies that ℓ_p divides $p-1$. If $p \equiv 2$ or $3 \pmod 5$ for an odd prime p, then $\left(\frac{5}{p}\right) = -1$ and $x^2 - x - 1$ is irreducible in $\mathbb{Z}_p[x]$. In this case, the order of $-r_2 = -1$ in $\mathbb{Z}_p^\times$ is 2, and so ℓ_p divides $2(p+1)$, but divides neither $p+1$ nor $p-1$. The period of the Fibonacci sequence modulo p can be smaller than the upper limits noted here for $p \neq 2, 5$. For example, $\ell_{29} = 14$, a proper divisor of $29-1$, and $\ell_{47} = 32$, a proper divisor of $2(47+1) = 96$ that divides neither 48 nor 46.

As we see here, unless the discriminant D of a quadratic polynomial f is identically zero, there are only finitely many primes p for which f has repeated factors in $\mathbb{Z}_p[x]$. The following generalization of the discriminant for higher degree polynomials similarly allows us (in theory) to determine all values of p for which a given polynomial f factors into distinct irreducible terms in $\mathbb{Z}_p[x]$. Let f be a monic polynomial of degree k with integer coefficients, which we can view as elements of $\mathbb{Z}$ or of $\mathbb{Z}_p$ for a prime p. Then f has k roots (not necessarily distinct) in some extension field of $\mathbb{Q}$ or $\mathbb{Z}_p$, and we can write

$$f(x) = x^k - r_1 x^{k-1} - \cdots - r_{k-1}x - r_k = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_k).$$

By definition, the *discriminant* of f is the product of the squares of all differences between the roots of f:

$$D = D(f) = \prod_{1 \le i < j \le k} (\alpha_j - \alpha_i)^2.$$

It immediately follows that $D(f) = 0$ if and only if f has a repeated root, that is, $\alpha_i = \alpha_j$ for some $i \neq j$. Note that D is a *symmetric polynomial* in $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$, meaning that it is unchanged by any permutation of the elements of that set. It is known that any such symmetric polynomial can be expressed in terms of *elementary* symmetric polynomials, which are, up to sign, the same as the coefficients of f. In general, if the coefficients of f are integers, then $D(f)$ is an integer which can be expressed in terms of those coefficients. (See [Edwards 1984] or [Swan 1962] for more details on computation of D.)

6. Criteria for factorization of polynomials modulo primes

Let f be a monic polynomial with integer coefficients, having degree k and discriminant D. In this section, we restrict our attention to primes p for which $p \nmid D$, so that f has no repeated irreducible factors in $\mathbb{Z}_p[x]$. We say that f has *factorization type* $[k_1, k_2, \ldots, k_j]$ modulo p if f can be written in $\mathbb{Z}_p[x]$ as a product of distinct irreducible polynomials having degrees $k_1 \ge k_2 \ge \cdots \ge k_j$. The number of possible factorization types of a polynomial of degree k is the number of *partitions* of k, that is, the number of ways of writing k as a sum of positive integers.

Theorem 9 implies that if we know the factorization type of a polynomial f modulo a prime p that divides neither $D(f)$ nor the constant coefficient of f, then we can use the order of the automorphism σ_p of $\mathbb{Z}_p[x]/\langle f\rangle$ to obtain information about the period length of a corresponding recursive sequence modulo p. We show in this section that we can reverse this implication for polynomials f of degree $k \leq 5$, using Theorem 9 together with the following application of the discriminant due to Stickelberger, adapted from [Driver et al. 2005] and [Swan 1962].

Stickelberger's parity theorem. Let f be a monic polynomial of degree k in $\mathbb{Z}[x]$ and let p be a prime number not dividing the discriminant D of f. Suppose that f factors as a product of j distinct irreducible polynomials in $\mathbb{Z}_p[x]$. If p is odd, then $\left(\frac{D}{p}\right) = (-1)^{k-j}$, while if $p = 2$, then $D \equiv 5^{k-j} \pmod 8$.

Before stating our main theorem for this section, we illustrate, with an example, how knowledge of the period of a recursive sequence modulo p can help determine the factorization type of its characteristic polynomial modulo p.

Example. Define a_n for $n \geq 0$ by $(a_0, a_1, a_2, a_3) = (0, 0, 0, 1)$ and $a_n = a_{n-3} + a_{n-4}$ for $n \geq 4$. The characteristic polynomial for $\{a_n\}_{n=0}^{\infty}$ is

$$f(x) = x^4 - x - 1,$$

which can be shown to have discriminant $D = -283$. So f is a product of distinct irreducible polynomials in $\mathbb{Z}_p[x]$ for all primes $p \neq 283$. Suppose that we calculate that modulo $p = 61$, the sequence $\{a_n\}_{n=0}^{\infty}$ has period $\ell = 75660$, which must be the same as the order of ω as a unit in $\mathbb{Z}_{61}[\omega] = \mathbb{Z}_{61}[x]/\langle f\rangle$. We find that ℓ divides neither $p - 1$ nor $p^2 - 1$, but does divide $p^3 - 1$. Theorem 9 implies that $t = 3$ is the least common multiple of the degrees of the irreducible factors of f in $\mathbb{Z}_{61}[x]$, and we conclude that f must have factorization type $[3, 1]$. Modulo $p = 71$, the same sequence has period $\ell = 1008$. This time we find that ℓ does not divide $p - 1$, but does divide $p^2 - 1$. Now f could have factorization type either $[2, 2]$ or $[2, 1, 1]$. But since $\left(\frac{-283}{71}\right) = 1$, Stickelberger's theorem implies that the number of irreducible factors of f in $\mathbb{Z}_{71}[x]$ has the same parity as $k = 4$, and so f has factorization type $[2, 2]$.

Remark. For computational purposes in this application, we can bypass direct calculation of the period of a recursive sequence. As noted in the example, this period ℓ is the same as the order of ω as a unit in a corresponding ring $\mathbb{Z}_p[\omega]$, so that ℓ divides an integer n precisely when $\omega^n = 1$. Powers of ω can be computed very efficiently by the process of *successive squaring*. If we write n in its binary expansion as

$$n = c_0 + c_1 \cdot 2 + c_2 \cdot 2^2 + c_3 \cdot 2^3 + \cdots ,$$

where $c_i = 0$ or 1 for all i, with only finitely many nonzero values of c_i, then

$$\omega^n = \omega^{c_0} \cdot (\omega^2)^{c_1} \cdot (\omega^4)^{c_2} \cdot (\omega^8)^{c_3} \cdots .$$

Each power of ω in parentheses is the square of the preceding power of ω, and only those values for which $c_i = 1$ contribute to the product. Squares and other products in

$$\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$$

are easily calculated by multiplying polynomials, replacing products by their remainders on division by f, when necessary.

Our next theorem states that we can determine the factorization type of a polynomial f of degree $k \le 5$ modulo most primes p (assuming that neither the discriminant nor the constant coefficient of f is identically zero) from knowledge of the discriminant of f and calculation of certain powers of ω in the ring $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$.

Theorem 11. *Let f be a monic polynomial with integer coefficients, having degree $k \le 5$ and discriminant D. Let p be a prime number that divides neither D nor the constant coefficient of f. Let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f \rangle$, and let t be the smallest positive integer such that $\omega^{p^t-1} = 1$ in $\mathbb{Z}_p[\omega]$. Then the following statements are true about the factorization of f in the ring $\mathbb{Z}_p[x]$.*

(1) *If $t = 1$, then f is a product of k distinct linear polynomials.*

(2) *If $t = 2$, and p is odd and $\left(\frac{D}{p}\right) = 1$, then f is a product of two distinct irreducible quadratic polynomials and $k - 4$ linear polynomials.*

(3) *If $t = 2$, and p is odd and $\left(\frac{D}{p}\right) = -1$, or $p = 2$ and $D \equiv 5 \pmod 8$, then f is a product of an irreducible quadratic polynomial and $k - 2$ distinct linear polynomials.*

(4) *If $t = 3$, then f is a product of an irreducible cubic polynomial and $k - 3$ distinct linear polynomials.*

(5) *If $t = 4$, then f is a product of an irreducible quartic polynomial and $k - 4$ linear polynomials.*

(6) *If $t = 5$, then f is an irreducible quintic polynomial.*

(7) *If $t = 6$, then f is a product of an irreducible cubic polynomial and an irreducible quadratic polynomial.*

Remark. As defined, the integer t is the same as the order of the automorphism σ_p in $\mathrm{Aut}(\mathbb{Z}_p[\omega])$, so must exist. It is understood that not all of the cases listed above can occur for every value of $k \le 5$, nor for every prime p. For example, case (2) is impossible when $p = 2$, since there are not two distinct irreducible quadratic polynomials in $\mathbb{Z}_2[x]$.

Proof. The table lists the seven partitions $[k_1, k_2, \dots, k_j]$ of $k = 5$.

$[k_1, k_2, \dots, k_j]$	$(-1)^{k-j}$	$t = \operatorname{lcm}(k_1, k_2, \dots, k_j)$
[1, 1, 1, 1, 1]	1	1
[2, 1, 1, 1]	−1	2
[3, 1, 1]	1	3
[2, 2, 1]	1	2
[4, 1]	−1	4
[3, 2]	−1	6
[5]	1	5

In the second column of the table, we note the parity of $k - j$ by listing $(-1)^{k-j}$, and in the third column, we list the least common multiple of the summands of the partition, which we label as t. Theorem 9 implies that if a polynomial f of degree five has factorization type $[k_1, k_2, \dots, k_j]$, then t is the smallest positive integer for which $\omega^{p^t-1} = 1$ in $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$, as in the statement of Theorem 11. The table shows that $t \le 6$, and that if $t \ne 2$, the factorization type of f is determined by the value of t. If $t = 2$, the factorization type of f is determined by t together with the value of $\left(\frac{D}{p}\right) = (-1)^{k-j}$ or $D \equiv 5^{k-j} \pmod 8$.

Removal of a term of 1, from those partitions containing 1, affects neither $(-1)^{k-j}$ nor t. (If a 1 is removed, both k and j are decreased by one, so that the value of $k - j$ is unchanged.) So the first five rows of the table lead to the same conclusion about polynomials of degree four; the first three rows imply the same about polynomials of degree three; and so forth. □

We now state three corollaries of Theorem 11, which can be viewed as algorithms for determining the factorization types of cubic, quartic, and quintic polynomials modulo prime values. Here we take better advantage of the Legendre symbol $\left(\frac{D}{p}\right)$, which is easy to calculate for a given D and odd prime p, as a first test to distinguish between factorization types. We omit the proofs, which follow the same arguments from the table exhibited in the proof of Theorem 11.

Corollary 12. *Let f be a monic polynomial of degree three with discriminant D, let p be a prime number that divides neither D nor the constant coefficient of f, and let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$.*

- *If p is odd and $\left(\frac{D}{p}\right) = 1$ or $p = 2$ and $p \equiv 1 \pmod 8$, then*:

(1) *If $\omega^{p-1} = 1$, then f has factorization type* [1, 1, 1].

(2) *If $\omega^{p-1} \ne 1$, then f has factorization type* [3].

- *If p is odd and $\left(\frac{D}{p}\right) = -1$ or $p = 2$ and $p \equiv 5 \pmod 8$, then*:

(3) *f has factorization type* [2, 1].

Corollary 13. *Let f be a monic polynomial of degree four with discriminant D, let p be a prime number that divides neither D nor the constant coefficient of f, and let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$.*

- *If p is odd and $\left(\frac{D}{p}\right) = 1$ or $p = 2$ and $p \equiv 1 \pmod 8$, then*:

(1) *If $\omega^{p-1} = 1$, then f has factorization type* $[1, 1, 1, 1]$.

(2) *If $\omega^{p-1} \neq 1$, but $\omega^{p^2-1} = 1$, then f has factorization type* $[2, 2]$.

(3) *If $\omega^{p^2-1} \neq 1$, then f has factorization type* $[3, 1]$.

- *If p is odd and $\left(\frac{D}{p}\right) = -1$ or $p = 2$ and $p \equiv 5 \pmod 8$, then*:

(4) *If $\omega^{p^2-1} = 1$, then f has factorization type* $[2, 1, 1]$.

(5) *If $\omega^{p^2-1} \neq 1$, then f has factorization type* $[4]$.

Corollary 14. *Let f be a monic polynomial of degree five with discriminant D, let p be a prime number that divides neither D nor the constant coefficient of f, and let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$.*

- *If p is odd and $\left(\frac{D}{p}\right) = 1$ or $p = 2$ and $p \equiv 1 \pmod 8$, then*:

(1) *If $\omega^{p-1} = 1$, then f has factorization type* $[1, 1, 1, 1, 1]$.

(2) *If $\omega^{p-1} \neq 1$, but $\omega^{p^2-1} = 1$, then f has factorization type* $[2, 2, 1]$.

(3) *If $\omega^{p^2-1} \neq 1$, but $\omega^{p^3-1} = 1$, then f has factorization type* $[3, 1, 1]$.

(4) *If $\omega^{p^2-1} \neq 1$ and $\omega^{p^3-1} \neq 1$, then f has factorization type* $[5]$.

- *If p is odd and $\left(\frac{D}{p}\right) = -1$ or $p = 2$ and $p \equiv 5 \pmod 8$, then*:

(5) *If $\omega^{p^2-1} = 1$, then f has factorization type* $[2, 1, 1, 1]$.

(6) *If $\omega^{p^2-1} \neq 1$, but $\omega^{p^4-1} = 1$, then f has factorization type* $[4, 1]$.

(7) *If $\omega^{p^4-1} \neq 1$, then f has factorization type* $[3, 2]$.

Remark. If t is the order of σ_p in the group of automorphisms of $\mathbb{Z}_p[\omega]$, then $\omega^{p^s-1} = 1$ if and only if t divides s. For example, in case (7) of Corollary 14, if $\omega^{p^4-1} \neq 1$, we are also claiming that $\omega^{p^2-1} \neq 1$.

Remark. As an example to illustrate the efficiency of these algorithms, a computer program written by the first author, based on Corollary 13, found the factorization type of $f(x) = x^4 - x - 1$ modulo all primes $p < 10000$ ($p \neq 283$) in approximately two seconds. On the same computer, a program to factor f in $\mathbb{Z}_p[x]$ for the same primes p, using brute force calculations, required four hours and 42 minutes to run. (The second program confirmed all of the results predicted by the first program.)

Polynomials of degree $k > 5$ cannot be distinguished from each other, in every case, by the same data. For example, if a polynomial f of degree six satisfies

$$\left(\frac{D(f)}{p}\right) = -1 \quad \text{and} \quad (\omega_f)^{p-1} \neq 1 \quad \text{but} \quad (\omega_f)^{p^2-1} = 1,$$

then f could have factorization type either $[2, 2, 2]$ or $[2, 1, 1, 1, 1]$. We conclude, however, with some results that hold for any value of k.

Theorem 15. *Let f be a monic polynomial of degree k with discriminant D, let p be a prime number that divides neither D nor the constant coefficient of f, and let $\mathbb{Z}_p[\omega] = \mathbb{Z}_p[x]/\langle f\rangle$.*

(1) *If $\omega^{p-1} = 1$, then f is a product of k linear factors in $\mathbb{Z}_p[x]$.*

(2) *If $\omega^{p^2-1} = 1$, then all irreducible factors of f in $\mathbb{Z}_p[x]$ have degree one or two. The number of irreducible quadratic factors of f is even if and only if p is odd and $\left(\frac{D}{p}\right) = 1$ or $p = 2$ and $D \equiv 1 \pmod 8$.*

(3) *If $\omega^{p^q-1} = 1$ for some odd prime q, then all irreducible factors of f in $\mathbb{Z}_p[x]$ have degree one or q. This case can occur only when p is odd and $\left(\frac{D}{p}\right) = 1$ or $p = 2$ and $D \equiv 1 \pmod 8$.*

Proof. Let the factorization type of f modulo p be $[k_1, k_2, \ldots, k_j]$, and let $t = \operatorname{lcm}(k_1, k_2, \ldots, k_j)$. If $\omega^{p-1} = 1$, then $t = 1$, which is possible only when $k_i = 1$ for $1 \le i \le j$, so that $j = k$. If $\omega^{p^q-1} = 1$ for some prime q, then t divides q. This is possible only when there is some $0 \le \ell \le j$ so that $k_i = q$ for $i \le \ell$ and $k_i = 1$ for $\ell < i \le j$. (We allow the possibility that $\ell = 0$, so that $t = 1$.) In this case, notice that $k = \ell \cdot q + (j - \ell)$, so that $k - j = \ell(q - 1)$. If $q = 2$, then $k - j$ has the same parity as ℓ. If q is odd, then $k - j$ is even in every case. □

Acknowledgement

The authors thank the Jepson Summer Science Institute at the University of Mary Washington for its financial support during the research project that led to this article.

References

[Adams 1984] W. W. Adams, "Splitting of quartic polynomials", *Math. Comp.* **43**:167 (1984), 329–343. MR 85f:12005 Zbl 0551.12017

[Driver et al. 2005] E. Driver, P. A. Leonard, and K. S. Williams, "Irreducible quartic polynomials with factorizations modulo p", *Amer. Math. Monthly* **112**:10 (2005), 876–890. MR 2006f:11027

[Dummit and Foote 2004] D. S. Dummit and R. M. Foote, *Abstract algebra*, 3rd ed., Wiley, Hoboken, NJ, 2004. MR 2007h:00003 Zbl 1037.00003

[Edwards 1984] H. M. Edwards, *Galois theory*, Graduate Texts in Mathematics **101**, Springer, New York, 1984. MR 87i:12002 Zbl 0532.12001

[Engstrom 1931] H. T. Engstrom, "On sequences defined by linear recurrence relations", *Trans. Amer. Math. Soc.* **33**:1 (1931), 210–218. MR 1501585 Zbl 0001.14002

[Fillmore and Marx 1968] J. P. Fillmore and M. L. Marx, "Linear recursive sequences", *SIAM Rev.* **10** (1968), 342–353. MR 38 #1936 Zbl 0169.51004

[Hardy and Wright 1979] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., Oxford University Press, New York, 1979. MR 81i:10002 Zbl 0423.10001

[Lidl and Niederreiter 1983] R. Lidl and H. Niederreiter, *Finite fields*, Encyclopedia of Mathematics and its Applications **20**, Addison-Wesley, Reading, MA, 1983. MR 86c:11106 Zbl 0554.12010

[Skolem 1952] T. Skolem, "The general congruence of the 4th degree modulo p, p prime", *Norsk Mat. Tidsskr.* **34** (1952), 73–80. MR 14,353e Zbl 0048.02905

[Sun 2003] Z.-H. Sun, "Cubic and quartic congruences modulo a prime", *J. Number Theory* **102**:1 (2003), 41–89. MR 2004e:11004 Zbl 1033.11003

[Sun 2006] Z.-H. Sun, "A criterion for polynomials to be congruent to the product of linear polynomials $\pmod p$", *Fibonacci Quart.* **44**:4 (2006), 326–329. MR 2008c:11005 Zbl 1160.11302

[Sun and Sun 1992] Z. H. Sun and Z. W. Sun, "Fibonacci numbers and Fermat's last theorem", *Acta Arith.* **60**:4 (1992), 371–388. MR 93e:11025 Zbl 0725.11009

[Swan 1962] R. G. Swan, "Factorization of polynomials over finite fields", *Pacific J. Math.* **12** (1962), 1099–1106. MR 26 #2432 Zbl 0113.01701

[Wall 1960] D. D. Wall, "Fibonacci series modulo m", *Amer. Math. Monthly* **67** (1960), 525–532. MR 22 #10945 Zbl 0101.03201

[Ward 1933] M. Ward, "The arithmetical theory of linear recurring series", *Trans. Amer. Math. Soc.* **35**:3 (1933), 600–628. MR 1501705 Zbl 0007.24901

Received: 2007-10-29 Accepted: 2010-01-26

llehman@umw.edu
University of Mary Washington, Department of Mathematics,
1301 College Avenue, Fredericksburg, VA 22401,
United States

ctriola@mw.edu
9 Seneca Terrace, Fredericksburg, VA 22401, United States

The Gram determinant for plane curves

Józef H. Przytycki and Xiaoqi Zhu

(Communicated by Kenneth S. Berenhaut)

We investigate the Gram determinant of the pairing arising from curves in a planar surface, with a focus on the disk with two holes. We prove that the determinant based on $n-1$ curves divides the determinant based on n curves. Motivated by the work on Gram determinants based on curves in a disk and curves in an annulus (Temperley–Lieb algebra of type A and B, respectively), we calculate several examples of the Gram determinant based on curves in a disk with two holes, and advance conjectures on the complete factorization of Gram determinants.

1. Introduction

Gram matrices and Gram determinants. Let B be a finite set and R a commutative ring. A *pairing* over B is a map $B \times B \to R$, denoted by $\langle \cdot , \cdot \rangle$. A very simple case is the Kronecker delta,

$$\langle i, j \rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \qquad \text{for } i, j \in B.$$

Let $b_1, \dots, b_n$ be a list of the elements of B, with $b_i \neq b_j$ if $i \neq j$. The *Gram matrix* of the pairing $\langle \cdot , \cdot \rangle$ is the $n \times n$ matrix

$$G = [\langle b_i, b_j \rangle]_{1 \leq i, j \leq n},$$

and the *Gram determinant* is the determinant of this matrix.

The name is derived from the classical case where R is a field, $B = \{b_1, \dots, b_n\}$ is a set of points in a vector space V over R, and the pairing is given by an inner product $\langle \cdot , \cdot \rangle$ on V. This situation is familiar; for instance, B is an orthonormal basis of V if and only if V has dimension n and the pairing coincides with the Kronecker delta described above.

MSC2000: 05A99, 57M99.

Keywords: Gram determinants, planar curves, noncrossing partitions, chromatic joins, Temperley–Lieb.

The Gram determinant plays a significant role in the classical case; for example, a set of vectors $B = \{b_1, \dots, b_n\} \subset V$ is linearly independent if and only if the Gram determinant of B is nonzero.

In our situation B will be a certain set of equivalence classes arising from sets of curves on a disk with holes. The ring R is a polynomial ring in many variables, and the pairing describes the interaction between the sets of curves when two copies of the disk are glued along their outer boundaries.[1]

The Gram matrix for a system of plane curves. Let F_0^n be a unit disk with $2n$ points on its boundary. Let $\boldsymbol{B}_0^n$ be the set of all possible diagrams, up to deformation, in F_0^n with n noncrossing chords connecting these $2n$ points. It is known that $|\boldsymbol{B}_0^n|$ is equal to the n-th Catalan number $C_n := \binom{2n}{n}/(n+1)$; see [Stanley 1999], for example. Accordingly, we will call $\boldsymbol{B}_0^n$ the set of *Catalan states*.

Consider the following generalized setup. Let $F_k \subset D^2$ be a plane surface with $k+1$ boundary components, which are given distinct labels. In particular, $F_0 = D^2$, and for $k \geq 1$, F_k is equal to D^2 with k holes. Let F_k^n be F_k with $2n$ points, $a_0, \dots, a_{2n-1}$, arranged counterclockwise along the outer boundary; see Figure 1, left. Throughout this paper, we use a_k and a_{k-1} to denote two adjacent points along the outer boundary, where k is taken modulo $2n$.

Let $\boldsymbol{B}_k^n$ be the set of all possible diagrams, up to equivalence, in F_k^n with n noncrossing chords connecting these $2n$ points, where equivalence is defined as follows: for each diagram $b \in \boldsymbol{B}_k^n$, there is a corresponding diagram $\gamma(b) \in \boldsymbol{B}_0^n$ obtained by filling the k holes in b. We call $\gamma(b)$ the underlying Catalan state of b (see Figure 1, right).

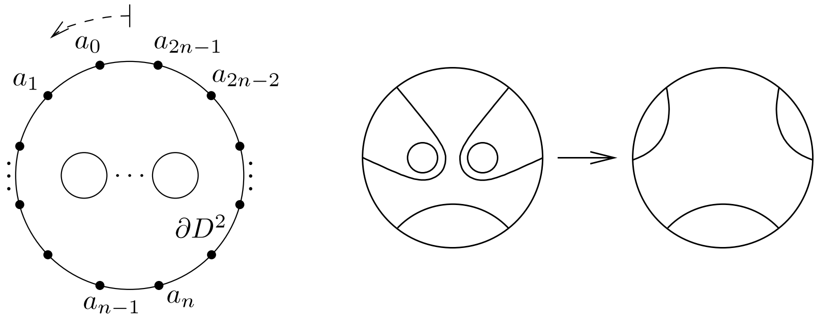


Figure 1. Left: notational conventions for F_k^n. Right: action of γ.

[1] A pairing over B extends to a bilinear form on the free R-module over B; this form is similar to the inner product on a vector space. Note, however, that the inner product over a *complex* vector space is linear in the first variable only, and conjugate-linear in the second; additionally, it is positive definite and conjugate-symmetric (skew-symmetric). The corresponding bilinear form in the more general setting need not be positive definite, symmetric, or conjugate-symmetric, although we will see echos of these properties in our situation.

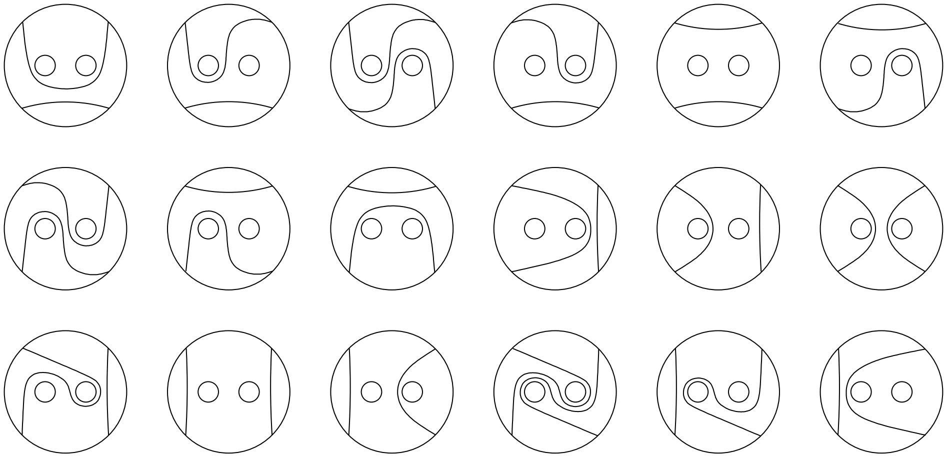


Figure 2. List of diagrams in $\boldsymbol{B}^2 = \boldsymbol{B}_2^2$.

A given diagram in F_k^n partitions F_k into $n+1$ regions. Two diagrams are equivalent if and only if they have the same underlying Catalan state and the labeled holes are distributed in the same manner across regions. Accordingly, $\boldsymbol{B}_k^n$ has elements $(n+1)^{k-1}\binom{2n}{n}$. See Figure 2 for the 18 diagrams in $\boldsymbol{B}_2^2$.

We remark that if $k=0$ or $k=1$, two diagrams are equivalent only if they are homotopic, but for $k>2$, this need not be true; see Figure 3 for a counterexample.

The study of noncrossing partitions of n points has a long history in enumerative combinatorics. Beyond purely combinatorial questions, noncrossing partitions arise in the study of a number of problems lying at the intersection of combinatorics and topology. Lickorish examines the matrix of a bilinear form defined on noncrossing planar diagrams in a disk, motivated by the theory of 3-manifold invariants. Motivated by the work of Birkhoff and Lewis [1946] on the four color conjecture, Tutte [1991] introduced the matrix of chromatic joins.

In this paper, we define a pairing over $\boldsymbol{B}_k^n$ and investigate the Gram matrix of the pairing. This concept is a generalization of a problem posed by W. B. R. Lickorish [1991; 1997] for type A Gram determinants — those based on a disk, i.e., $k=0$ —

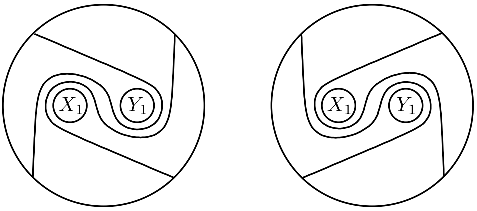


Figure 3. Two nonisotopic but equivalent diagrams in F_2^2; they correspond to the same state in $\boldsymbol{B}^2$.

and by Rodica Simion for type B Gram determinants [Schmidt 2004; Simion 2000] (these are related to $k = 1$ and the Kauffman bracket skein module of an annulus; see [Przytycki 1999]). Simion was motivated by Tutte's work [1991; 1993] on chromatic joins; see also [Chen and Przytycki 2008].

Significant research has been completed for the Gram determinants for type A and B. In particular, Di Francesco [1998] and Westbury [1995] gave a closed formula for the type A Gram determinant; a complete factorization of the type B Gram determinant was conjectured by Gefry Barad, and a closed formula (quoted in Theorem 7.4) was proved by Martin and Saleur [1993] and by Chen and Przytycki [2009]. The type A Gram determinant was used by Lickorish to find an elementary construction of Reshetikhin–Turaev–Witten invariants of oriented closed 3-manifolds.

We specifically investigate the Gram determinant G_n of the bilinear form defined over $\boldsymbol{B}_2^n$ and prove that $\det G_{n-1}$ divides $\det G_n$ for $n > 1$. Furthermore, we investigate the diagonal entries of G_n and give a method for computing terms of maximal degree in $\det G_n$. We conclude the paper by briefly discussing generalizations of the Gram determinant and presenting some open questions.

2. Definitions and basic facts for $\boldsymbol{B}_2^n$

Consider F_2^n, a unit disk with two holes, along with $2n$ points along the outer boundary. Denote the holes in F_2^n by X_1 and Y_1. To differentiate between them, we will always place X_1 to the left and Y_1 to the right if labels are not present.

Let

$$\boldsymbol{B}^n := \boldsymbol{B}_2^n =: \left\{b_1, \ldots, b_{(n+1)\binom{2n}{n}}\right\}$$

be the set of all possible diagrams with n noncrossing chords connecting these $2n$ points, up to equivalence in F_2^n.

Recall that in complex analysis an *inversion* (in the unit circle) is the involution defined on the sphere $S^2 = \mathbb{C} \cup \infty$ by $z \leftrightarrow z/|z|^2$. Let X_2 and Y_2 be the inversions of X_1 and Y_1, respectively, and let $\mathcal{S} = \{X_1, X_2, Y_1, Y_2\}$. Given $b_i \in \boldsymbol{B}^n$, let b_i^* denote the inversion of b_i. Given $b_i, b_j \in \boldsymbol{B}^n$, we glue b_i with b_j^* along the outer boundary, respecting the labels of the marked points. Since b_i and b_j each contains n noncrossing chords, $b_i \circ b_j^*$ can have at most n closed curves. The resulting diagram, denoted by $b_i \circ b_j^*$, is a set of up to n closed curves in the 2-sphere $S^2 = D^2 \cup (D^2)^*$ with four holes, X_1, X_2, Y_1, Y_2. (Since we glued along ∂D^2, it is no longer a boundary.) Each closed curve partitions the set $\mathcal{S}$ into two sets. Two closed curves are of the same type if they partition $\mathcal{S}$ the same way.

We define a pairing $\langle \cdot\,, \cdot \rangle$ over $\boldsymbol{B}^n$ by associating with $b_i, b_j \in \boldsymbol{B}^n$ a monomial in the variables $d, x_1, x_2, y_1, y_2, z_1, z_2, z_3$, as follows. The exponent of each variable is obtained by counting the number of curves in $b_i \circ b_j^*$ that partition the set $\{X_1, X_2, Y_1, Y_2\}$ in the corresponding way, the correspondence being this:

x_1 :	$\{X_1\}, \{X_2, Y_1, Y_2\}$	z_1:	$\{X_1, X_2\}, \{Y_1, Y_2\}$
x_2 :	$\{X_2\}, \{X_1, Y_1, Y_2\}$	z_2:	$\{X_1, Y_1\}, \{X_2, Y_2\}$
y_1 :	$\{Y_1\}, \{X_1, X_2, Y_2\}$	z_3:	$\{X_1, Y_2\}, \{X_2, Y_1\}$
y_2 :	$\{Y_2\}, \{X_1, X_2, Y_1\}$	d:	$\varnothing, \{X_1, X_2, Y_1, Y_2\}$

Table 1. Indeterminates and partitions. In the monomial $\langle b_i, b_j \rangle$, the exponent of each variable is the number of curves in $b_i \circ b_j^*$ that partition the set $\{X_1, X_2, Y_1, Y_2\}$ in the given way.

Thus $\langle b_i, b_j \rangle$ is a monomial of degree at most n. Some example paired diagrams, with their corresponding monomials, are given in Figure 5.

We can now form the Gram matrix $G_n = [g_{ij}] = [\langle b_i, b_j \rangle]_{1 \le i, j \le (n+1)\binom{2n}{n}}$ of this pairing. We write it explicitly for $n = 1$. Order the elements of $\boldsymbol{B}^1$ as in the first

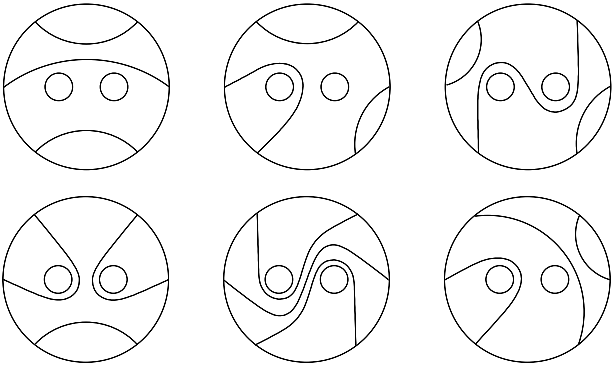
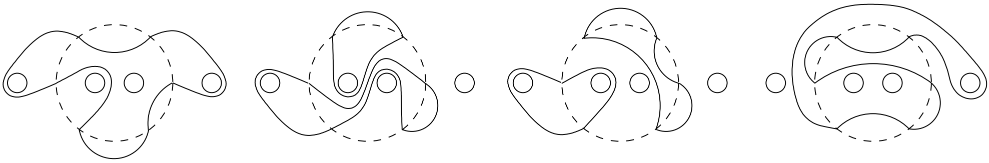


Figure 4. Diagrams of six states $b_1, b_2, b_3, b_4, b_5, b_6 \subset \boldsymbol{B}^3$. The indices are used in the examples, but are not intrinsic.



$\langle b_2, b_4 \rangle = x_1$ $\langle b_5, b_2 \rangle = x_1 x_2$ $\langle b_6, b_2 \rangle = d z_1$ $\langle b_1, b_3 \rangle = x_2$

Figure 5. Diagrams for $b_i \circ b_j$ on S^2 and the corresponding values of $\langle b_i, b_j \rangle$. Indices are as in Figure 4.

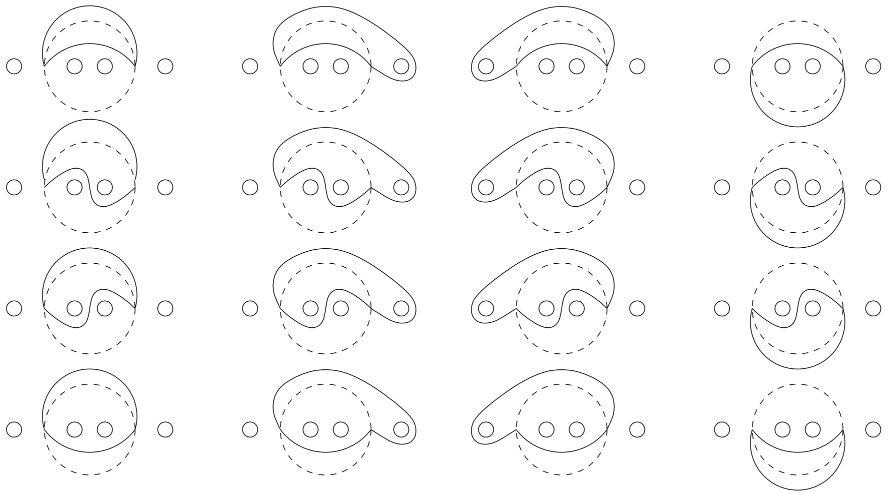


Figure 6. Array of $b_i \circ b_j^*$ for $b_i, b_j \in \boldsymbol{B}^1$.

column of Figure 6 (we're looking at the disk inside the dotted circle). Then we see from the array of diagrams in Figure 6, each of which represents one pair (b_i, b_j), that the Gram matrix of the pairing is

$$G_1 = \begin{bmatrix} d & y_2 & x_2 & z_2 \\ y_1 & z_1 & z_3 & x_1 \\ x_1 & z_3 & z_1 & y_1 \\ z_2 & x_2 & y_2 & d \end{bmatrix}.$$

Therefore the Gram determinant is

$$\det G_1 = (dz_1 - z_1z_2 - dz_3 + z_2z_3 - x_1x_2 + x_2y_1 + x_1y_2 - y_1y_2) \\ (dz_1 + z_1z_2 + dz_3 + z_2z_3 - x_1x_2 + x_2y_1 - x_1y_2 + y_1y_2).$$

This paper is mostly devoted to exploring possible factorizations of $\det G_n$, and is the first step toward computing $\det G_n$ in full generality, which we conjecture to have a nice decomposition.

Though the pairing (and hence the Gram matrix) is not symmetric, it is skew-symmetric with respect to an certain involution of the ring R. (An involution is isomorphism equal to its own inverse.) Specifically, given $b_i, b_j \in \boldsymbol{B}^n$, we can obtain $b_j \circ b_i^*$ from $b_i \circ b_j^*$ by inversion in the unit circle, which interchanges X_1 with X_2 and Y_1 with Y_2. Consequently, $\langle b_j, b_i \rangle$ can be obtained from $\langle b_i, b_j \rangle$ by interchanging x_1 with x_2 and y_1 with y_2, as these interchanges have the same effect in the corresponding partition (see Table 1) as the hole interchange $X_1 \leftrightarrow X_2$, $Y_1 \leftrightarrow Y_2$. Note that z_1, z_2, z_3, and d are mapped to themselves under this variable

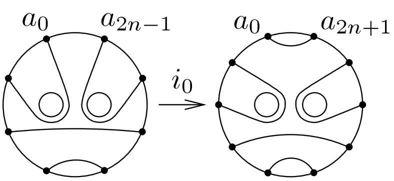


Figure 7. Action of the embedding i_0. Note the relabeling of the boundary points: each a_k on the left becomes a_{k+1} on the right, and the two new points are labeled a_0 and a_{2n+1}.

swap, because their partitions are invariant under the interchange of holes. (See also Theorem 3.3(4) below.)

To summarize, let h_t be the involution of G_n that interchanges x_1 with x_2 and y_1 with y_2. Then

$$\langle b_i, b_j \rangle = h_t(\langle b_j, b_i \rangle),$$

and the transpose of G_n is given by applying h_t to each individual entry of G_n.

Embedding B^n in B^{n+1}. Let $i_0 : \boldsymbol{B}^n \to \boldsymbol{B}^{n+1}$ be the *embedding* (injection) defined as follows: for $b_i \in \boldsymbol{B}^n$, the image $i_0(b_i) \in \boldsymbol{B}^{n+1}$ is given by adding to b_i a noncrossing chord close to the outer boundary and joining two points between a_0 and a_{2n-1}, as suggested in Figure 7. The two new points on the edge become the new a_0 and a_{2n+1}, and each of the old points a_k becomes a_{k+1}. This relabeling explicitly makes $i_0(b_i)$ an element of $\boldsymbol{B}^{n+1}$.

Another embedding we will need, denoted by $i_1 : \boldsymbol{B}^n \to \boldsymbol{B}^{n+1}$ and illustrated in Figure 8, is defined by a construction similar to that of i_0, but this time the added chord joins two points between the old a_0 and a_1, rather than between a_0 and a_{2n-1}. These two new points become a_0 and a_1, while the old a_0 becomes a_{2n+1} and each a_k, for $1 < k < 2n$, becomes a_{k+1}.

More formally, we define i_1 in terms of i_0 by using the notion of a *Dehn twist*, borrowed from surface topology and knot theory. Fix an annulus in the complex plane — the region between two concentric circles, say $R' \leq |z| \leq 1$. Imagine keeping the inner boundary circle fixed, while the outer one is rotated clockwise by an angle α. The stuff in between also gets rotated, by an amount that depends on how far it is from each circle. The resulting homeomorphism of the annulus is

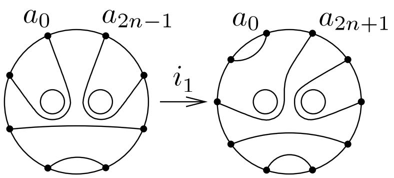


Figure 8. Action of the embedding i_1.

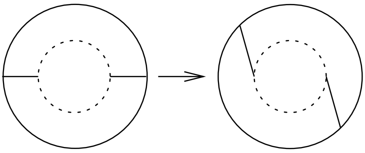


Figure 9. A Dehn twist r_α, with $\alpha = \pi/4$.

called a *Dehn twist* through an angle α. As an explicit formula we can take

$$r_\alpha(z) = z \exp\left(i\alpha \frac{|z| - R'}{1 - R'}\right),$$

which says the amount of rotation experienced by a point is proportional to the distance to the inner circle, growing from 0 at $|z| = R'$ to the full angle α at $|z| = 1$. Figure 9 gives a qualitative picture in the case $\alpha = \pi/4$.

Now we get back to the disk with two holes, F_2^n. If we choose R' close enough to 1 that the holes X_1 and Y_1 lie within the circle of radius R', we can extend r_α to a homeomorphism of F_2^n by setting $r_\alpha(z) = z$ for $|z| \le R'$.

Moreover, if $\alpha = \pi/n$, then r_α takes each of the $2n$ marked points a_k on the edge of F_2^n to the next such point a_{k+1}; consequently, it takes a system of noncrossing curves in F_2^n to another such. This defines the action of $r_{\pi/n}$ on $\boldsymbol{B}^n$; it is a permutation because the inverse of a Dehn twist is also a Dehn twist through the opposite angle.[2] The first arrow in Figure 10 illustrates the action of ${r_{\pi/4}}^{-1}$ on a certain element of $\boldsymbol{B}^4$, and the last arrow shows the action of $r_{\pi/5}$ on an element of $\boldsymbol{B}^5$.

We can now express i_1 in terms of i_0 and Dehn twists:

$$i_1 = r_{\pi/(n+1)} \circ i_0 \circ {r_{\pi/n}}^{-1}.$$

This is illustrated in Figure 10. Note that the two Dehn twists are not quite inverse to each other, since their angles differ.

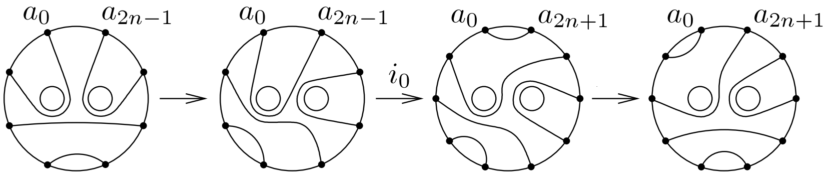


Figure 10. The embedding i_1, illustrated in Figure 8, is obtained from i_0 (Figure 7) by composing with appropriate Dehn twists.

[2] Obviously the repeated application of k Dehn twists through α is a Dehn twist by $k\alpha$, so any $r_{k\pi/n}$ also induces an action on $\boldsymbol{B}^n$. Note that the Dehn twist by a full 2π, though it is not the identity homeomorphism, gives the identity map on $\boldsymbol{B}^n$; an example of its action was shown in Figure 3.

3. More properties of the Gram determinant

Theorem 3.1. $\det G_n \neq 0$ *for all integers* $n \geq 1$.

Lemma 3.2. $\langle b_i, b_j \rangle$ *is a monomial of maximal degree if and only if* $\gamma(b_i) = \gamma(b_j)$.

Proof. Recall that $\langle b_i, b_j \rangle$ has maximal degree if and only if $b_i \circ b_j^*$ has n closed curves; this in turn is equivalent to having each closed curve made of exactly two arcs, one in b_i and one in b_j^*. In this situation, any two points connected by a chord in b_i must also be connected by a chord in b_j, so $\gamma(b_i) = \gamma(b_j)$. □

Proof of Theorem 3.1. Assume $\langle b_i, b_j \rangle$ is a monomial of maximal degree consisting only of the variables d and z_1. Because $\gamma(b_i) = \gamma(b_j)$ by Lemma 3.2, it follows that any two points connected in b_i are also connected in b_j. Each connection in b_i can be drawn in four different ways with respect to X and Y, since there are two ways to position the chord relative to each hole. Because $\langle b_i, b_j \rangle$ is assumed to consist only of the variables d and z_1, it follows that each pair of arcs that form a closed curve in $b_i \circ b_j^*$ either separates $\{X_1, X_2\}$ from $\{Y_1, Y_2\}$ or has $\{X_1, X_2, Y_1, Y_2\}$ on the same side of the curve. One can check each of the four cases to see that this condition implies that any two arcs that form a closed curve in $b_i \circ b_j^*$ must be equal, so $b_i = b_j$. Using Laplacian expansion, this implies that the product of the diagonal of G_n is the unique summand of degree $n(n+1)\binom{2n}{n}$ in $\det G_n$ consisting only of the variables d and z_1. □

We need the following notation for the next theorem: let $f : \alpha_1 \leftrightarrow \alpha_2$ denote a function f which acts on the entries of G_n by interchanging variables α_1 with α_2. We can extend the domain of f to G_n. Let $f(G_n)$ denote the matrix formed by applying f to all the individual entries of G_n.

Define involutions h_1, h_2, h_3, h_t acting on the entries of G_n as follows:

$$\begin{array}{lll} h_1 : & x_1 \leftrightarrow y_1 & z_1 \leftrightarrow z_3 \\ h_2 : & x_2 \leftrightarrow y_2 & z_1 \leftrightarrow z_3 \\ h_3 = h_1 h_2 : & x_1 \leftrightarrow y_1 & x_2 \leftrightarrow y_2 \\ h_t : & x_1 \leftrightarrow x_2 & y_1 \leftrightarrow y_2 \end{array}$$

Theorem 3.3.

(1) $\det h_1(G_1) = -\det G_1$, *and for* $n > 1$, $\det h_1(G_n) = \det G_n$.

(2) $\det h_2(G_1) = -\det G_1$, *and for* $n > 1$, $\det h_2(G_n) = \det G_n$.

(3) $\det h_3(G_n) = \det G_n$.

(4) $\det h_t(G_n) = \det G_n$.

Proof. For assertion (1), note that $h_1(G_n)$ corresponds to exchanging the positions of the holes X_1 and Y_1 for all $b_i \in \boldsymbol{B}^n$. b_j^* is unchanged, so h_1 can be realized by a permutation of rows. For states where X_1 and Y_1 lie in the same region, their corresponding rows are unchanged by h_1. The number of such states is given by $|\boldsymbol{B}^n|/(n+1)$. Thus, the total number of row transpositions is equal to

$$\frac{1}{2}\left(|\boldsymbol{B}^n| - \frac{|\boldsymbol{B}^n|}{n+1}\right) = \frac{n}{2}\binom{2n}{n} = \frac{n(n+1)}{2}C_n.$$

It is known that C_n is odd if and only if $n = 2^m - 1$ for some m; see for instance [Deutsch and Sagan 2006]. Hence, C_n being odd implies that

$$\frac{n(n+1)}{2} = \frac{2^m(2^m-1)}{2} = 2^{m-1}(2^m-1),$$

which is even for all $m > 1$. Thus, $h_1(G_n)$ can be obtained from G_n by an even permutation of rows for $n > 1$, so $\det h_1(G_n) = \det G_n$. Similarly, $h_1(G_1)$ is given by an odd number of row transpositions on G_1, so $\det h_1(G_1) = -\det G_1$.

Assertion (2) can be shown using the same argument, except that h_2 corresponds to interchanging the positions of the holes X_2 and Y_2, rather than X_1 and Y_1.

Since $h_3 = h_1 h_2$, it follows immediately that $\det h_3(G_n) = \det G_n$ for $n > 1$. The sum of two odd permutations is even, so the equality also holds for $n = 1$, which proves (3). Assertion (4) follows because $\det h_t(G_n) = \det {}^tG_n = \det G_n$. □

Theorem 3.4. $\det G_n$ *is preserved under the following involutions*:

$$\begin{array}{lllll}
g_1: & x_1 \leftrightarrow -x_1 & x_2 \leftrightarrow -x_2 & z_2 \leftrightarrow -z_2 & z_3 \leftrightarrow -z_3 \\
g_2: & y_1 \leftrightarrow -y_1 & y_2 \leftrightarrow -y_2 & z_2 \leftrightarrow -z_2 & z_3 \leftrightarrow -z_3 \\
g_3: & x_1 \leftrightarrow -x_1 & y_2 \leftrightarrow -y_2 & z_1 \leftrightarrow -z_1 & z_2 \leftrightarrow -z_2 \\
g_1g_2: & x_1 \leftrightarrow -x_1 & x_2 \leftrightarrow -x_2 & y_1 \leftrightarrow -y_1 & y_2 \leftrightarrow -y_2 \\
g_1g_3: & x_2 \leftrightarrow -x_2 & y_2 \leftrightarrow -y_2 & z_1 \leftrightarrow -z_1 & z_3 \leftrightarrow -z_3 \\
g_2g_3: & x_1 \leftrightarrow -x_1 & y_1 \leftrightarrow -y_1 & z_1 \leftrightarrow -z_1 & z_3 \leftrightarrow -z_3 \\
g_1g_2g_3: & x_2 \leftrightarrow -x_2 & y_1 \leftrightarrow -y_1 & z_1 \leftrightarrow -z_1 & z_2 \leftrightarrow -z_2
\end{array}$$

Proof. We first show that g_1 can be realized by conjugating the matrix G_n by a diagonal matrix P_n of all diagonal entries equal to ±1. Define the diagonal entries of P_n by

$$p_{ii} = (-1)^{q(b_i, F_x)},$$

where $q(b_i, F_x)$ is the number of times b_i intersects F_x modulo 2; see Figure 11, where F_x, F_x^*, F_y, F_y^* and $\tilde{F}_x$ are defined. F_x and F_y touch the unit circle between a_0 and a_{2n-1}.

This proves the result about g_1, because curves corresponding to the variables x_1, x_2, z_2 and z_3 intersect $F_x \cup {F_x}^*$ in an odd number of points, whereas curves corresponding to the variables d, z_2, y_1 and y_2 cut it an even number of times. More precisely, for

$$g_{ij} = \langle b_i, b_j\rangle = d^{n_d}{x_1}^{n_{x_1}}{x_2}^{n_{x_2}}{y_1}^{n_{y_1}}{y_2}^{n_{y_2}}{z_1}^{n_{z_1}}{z_2}^{n_{z_2}}{z_3}^{n_{z_3}},$$

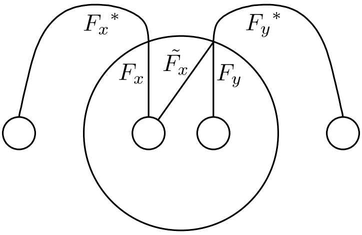


Figure 11. Toward the proof of Theorem 3.4(1).

the entry g'_{ij} of $P_n G_n {P_n}^{-1}$ satisfies

$$\begin{aligned} g'_{ij} &= p_{ii} g_{ij} p_{jj} = p_{ii} p_{jj} g_{ij} = (-1)^{q(b_i,F_x)+q(b_j,F_x)} g_{ij} = (-1)^{n_{x_1}+n_{x_2}+n_{z_2}+n_{z_3}} g_{ij} \\ &= d^{n_d}(-x_1)^{n_{x_1}}(-x_2)^{n_{x_2}} {y_1}^{n_{y_1}} {y_2}^{n_{y_2}} {z_1}^{n_{z_1}} (-z_2)^{n_{z_2}} (-z_3)^{n_{z_3}}. \end{aligned}$$

The results about g_2 and g_3 follow by the same argument, but using F_y and $F_y \cup {F_y}^*$ for g_2 and $\tilde{F}_x$ and $\tilde{F}_x \cup {F_y}^*$ for g_3. The statements about compositions follow directly from the first three. □

4. Terms of maximal degree in det G_n

Theorem 3.1 proves that the product of the diagonal entries of G_n is the unique term of maximal degree, $n(n+1)\binom{2n}{n}$, in $\det G_n$ consisting only of the variables d and z_1. More precisely, the product of the diagonal of G_n is given by

$$\delta(n) = \prod_{b_i \in \boldsymbol{B}^n} \langle b_i, b_i \rangle = d^{\alpha(n)} z_1^{\beta(n)},$$

with $\alpha(n) + \beta(n) = n(n+1)\binom{2n}{n}$. The value of $\delta(n)$ for the first few n are

$$\delta(1) = d^2 z_1^2, \quad \delta(2) = d^{20} z_1^{16}, \quad \delta(3) = d^{144} z_1^{96}, \quad \delta(4) = d^{888} z_1^{512}.$$

Computing the general formula for $\delta(n)$ can be reduced to a purely combinatorial problem. We conjectured that $\beta(n) = (2n)4^{n-1}$ and this was proven by Louis Shapiro (personal communication, 2008) using an involved generating function argument. The result is stated formally below.

Theorem 4.1. $\delta(n) = d^{n(n+1)\binom{2n}{n}-(2n)4^{n-1}} z_1^{(2n)4^{n-1}}$.

Let $h(\det G_n)$ denote the truncation of $\det G_n$ to terms of maximal degree, that is, of degree $n(n+1)\binom{2n}{n}$. Each term is a product of $(n+1)\binom{2n}{n}$ entries in G_n, each of which is a monomial of degree n. By Lemma 3.2, $\langle b_i, b_j \rangle$ has degree n if and only if b_i and b_j have the same underlying Catalan state. Divide $\boldsymbol{B}^n$ into subsets corresponding to underlying Catalan states, that is, into subsets $A_1, \dots, A_{C_n}$, such that for all $b_i, b_j \in A_k$, $\gamma(b_i) = \gamma(b_j)$. Then from Lemma 3.2 we have:

Proposition 4.2. *For $1 \leq k \leq C_n$, let I_k be the set of indices such that $A_k = \{b_i\}_{i \in I_k}$, and let $\langle A_k, A_k \rangle$ be the submatrix of G_n whose rows and columns are indexed by I_k. Then*

$$h(\det G_n) = \prod_{k=1}^{C_n} \det \langle A_k, A_k \rangle.$$

Note that the $\langle A_k, A_k \rangle$ are simply blocks in G_n, and their determinants can be multiplied together to give the highest terms in $\det G_n$. Finding the terms of maximal degree in $\det G_n$ can give insight into the decomposition of $\det G_n$ for large n.

Example 4.3. $\boldsymbol{B}^1$ corresponds to the single Catalan state in $\boldsymbol{B}^1_0$. Thus, $\det G_1 = h(\det G_1)$, a homogeneous polynomial of degree 4 (given on page 154).

Example 4.4. We can divide $\boldsymbol{B}^2$ into two sets, corresponding to the two Catalan states in $\boldsymbol{B}^2_0$. Thus $h(\det G_2)$ can be found by computing two 9×9 block determinants. The two Catalan states in $\boldsymbol{B}^2_0$ are equivalent up to rotation, so the two block determinants are equal. Specifically, we have:

$$\begin{aligned} h(\det G_2) &= d^6(x_1x_2 + x_2y_1 + x_1y_2 + y_1y_2 - dz_1 - z_1z_2 - dz_3 - z_2z_3)^4 \\ &\qquad (-x_1x_2 + x_2y_1 + x_1y_2 - y_1y_2 + dz_1 - z_1z_2 - dz_3 + z_2z_3)^4 \\ &\qquad (-x_1x_2z_1 - y_1y_2z_1 + dz_1{}^2 + x_2y_1z_3 + x_1y_2z_3 - dz_3{}^2)^2 \\ &\qquad (-2x_1x_2y_1y_2 + dx_1x_2z_1 + dy_1y_2z_1 - d^2z_1{}^2 + dx_2y_1z_3 + dx_1y_2z_3 - d^2z_3{}^2)^2 \\ &= d^6 \det G_1{}^4(-x_1x_2z_1 - y_1y_2z_1 + dz_1{}^2 + x_2y_1z_3 + x_1y_2z_3 - dz_3{}^2)^2 \\ &\qquad (-2x_1x_2y_1y_2 + dx_1x_2z_1 + dy_1y_2z_1 - d^2z_1{}^2 + dx_2y_1z_3 + dx_1y_2z_3 - d^2z_3{}^2)^2. \end{aligned}$$

Example 4.5. $\boldsymbol{B}^3$ can be divided into five subsets, corresponding to the five Catalan states in $\boldsymbol{B}^3_0$. We can thus find $h(\det G_3)$ by computing the determinants of five blocks in $\boldsymbol{B}^3$. The determinant of each block gives a homogeneous polynomial of degree $240/5 = 48$. $\boldsymbol{B}^3_0$ forms two equivalence classes up to rotation, so there are only two unique block determinants. The result is

$$\begin{aligned} &h(\det G_3) \\ &= h(\det G_2)^6 \det G_1{}^{-9} d^{30} w^3 \bar{w}^3 \\ &= d^{66}(-x_1x_2+x_2y_1+x_1y_2-y_1y_2+dz_1-z_1z_2-dz_3+z_2z_3)^{15} \\ &\quad (-x_1x_2-x_2y_1-x_1y_2-y_1y_2+dz_1+z_1z_2+dz_3+z_2z_3)^{15} \\ &\quad (-x_1x_2z_1-y_1y_2z_1+dz_1^2+x_2y_1z_3+x_1y_2z_3-dz_3{}^2)^{12} \\ &\quad (2x_1x_2y_1y_2-dx_1x_2z_1-dy_1y_2z_1+d^2z_1{}^2-dx_2y_1z_3-dx_1y_2z_3+d^2z_3{}^2)^{12} \\ &\quad (x_1x_2y_1y_2z_1-dx_1x_2z_1{}^2-dy_1y_2z_1{}^2+d^2z_1{}^3-x_1x_2y_1y_2z_3+dx_2y_1z_3{}^2+dx_1y_2z_3{}^2-d^2z_3{}^3)^3 \\ &\quad (x_1x_2y_1y_2z_1-dx_1x_2z_1^2-dy_1y_2z_1^2+d^2z_1{}^3+x_1x_2y_1y_2z_3-dx_2y_1z_3{}^2-dx_1y_2z_3{}^2+d^2z_3{}^3)^3. \end{aligned}$$

5. $\det G_{n-1}$ divides $\det G_n$

We defined in Section 2 the embeddings $i_0, i_1 : \boldsymbol{B}^n \to \boldsymbol{B}^{n+1}$. We now introduce inverses of sort for these two maps.

Given $b_i \in \boldsymbol{B}^n$, imagine adding to b_i a noncrossing chord connecting a_0 and a_{2n-1} outside the circle, and then pushing this chord inside the circle, together with the points a_0 and a_{2n-1}; see Figure 12. With the removal of these two points from the boundary, we relabel the remaining ones so the old a_k becomes a_{k-1}, for $0 < k < 2n-1$. So now there are $2n-2$ marked points on the boundary; this establishes a projection $\boldsymbol{B}^n \to \boldsymbol{B}^{n-1}$, with one caveat soon to be discussed. We denote this projection by p_0.

The procedure we've described works fine so long as b_i does not include a chord joining a_0 and a_{2n-1}. Indeed, if a_0 and a_{2n-1} are connected respectively to a_j and a_k in b_i, the added exterior chord ends up, in $p_0(b_i)$, as part of a chord joining a_{j-1} to a_{k-1} (see again Figure 12). However, a problem arises when b_i has a chord from a_0 to a_{2n-1}. In this case, the procedure creates a closed curve inside the disc, coming from the two chords joining the old a_0 to a_{2n-1}, one internal and one external. One could imagine erasing this loop to obtain an element of $\boldsymbol{B}^{n-1}$, but the loop carries information — it may enclose an arbitrary subset of $\{X_1, Y_1\}$. So we keep it at present, and we make p_0 take values in the set $\overline{\boldsymbol{B}}^{n-1}$ of equivalence classes of diagrams in F_2^{n-1} consisting of $n-1$ chords joining marked points on the boundary together with an optional closed loop disjoint from the boundary.

These observations can be summarized as follows:

Lemma 5.1. *An element $b_i \in \boldsymbol{B}^n$ is taken under $p_0 : \boldsymbol{B}^n \to \overline{\boldsymbol{B}}^{n-1}$ to an element of $\boldsymbol{B}^{n-1}$ if and only if b_i contains no chord connecting a_0 and a_{2n-1}.*

A bit of experimentation will persuade the reader of the correctness of the next result — which, incidentally, justifies our decision to expand the range of p_0 to include diagrams with a loop.

Proposition 5.2. *For any $b_i \in \boldsymbol{B}^n$ and $b_j \in \boldsymbol{B}^{n-1}$, we have*

$$b_i \circ i_0(b_j)^* = p_0(b_i) \circ b_j^*,$$

where the equivalence relation implicit in this equality consists of isotopies of the four-holed sphere, not necessarily preserving the unit disk.

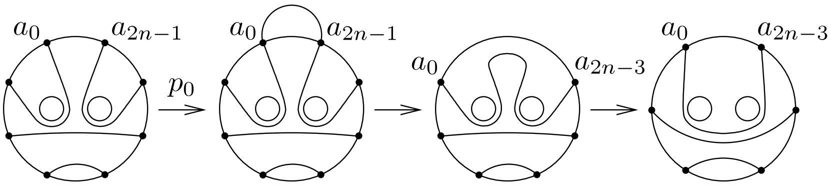


Figure 12. Action of the projection p_0.

We're gearing up toward a demonstration that the Gram determinant for $n-1$ chords divides the Gram determinant for n chords. We need one more lemma.

Lemma 5.3. *Fix $b_i \in \boldsymbol{B}^n$. There exists an element $b_{\alpha(i)} \in \boldsymbol{B}^{n-1}$ and a monomial $q \in \{1, d, x_1, y_1, z_2\}$ such that*

$$\langle p_0(b_i), b_j\rangle = q\,\langle b_{\alpha(i)}, b_j\rangle \quad \textit{for all } b_j \in \boldsymbol{B}^{n-1}.$$

Proof. If $p_0(b_i) \in \boldsymbol{B}^{n-1}$ we can take $b_{\alpha(i)} = p_0(b_i)$ and $q = 1$. Otherwise, it follows from Lemma 5.1 that b_i contains a chord connecting a_0 and a_{2n-1}, and $p_0(b_i)$ is the union of some $b_{\alpha(i)} \in \boldsymbol{B}^{n-1}$ with a loop enclosing a subset of $\{X_1, Y_1\}$. Let q be the variable corresponding to the partition of the holes effected by extra loop, according to Table 1. Then $\langle p_0(b_{\alpha(i)}), b_j\rangle = q\,\langle b_{\alpha(i)}, b_j\rangle$ for any b_j. □

For the remainder of the paper we adopt the following notation: if B and B' are subsets of $\boldsymbol{B}^n$, let

$$\langle B, B'\rangle := \big[\langle b_i, b_j\rangle\big]_{\substack{i\,:\,b_i\in B\\ j\,:\,b_j\in B'}}$$

be the submatrix of G_n whose rows correspond to the elements of B and whose columns correspond to the elements of B'.

Theorem 5.4. *For $n > 1$,* $\det G_{n-1}$ *divides* $\det G_n$.

Proof. We use the easily checked equality (also proved in detail as Lemma 6.1)

$$\langle i_0(b_i), i_1(b_j)\rangle = \langle i_1(b_i), i_0(b_j)\rangle = \langle b_i, b_j\rangle \quad \text{for all } b_i, b_j \in \boldsymbol{B}^{n-1}.$$

In the notation defined before the theorem, this means that $\langle b_i, \boldsymbol{B}^{n-1}\rangle$ (the i-th row of G_{n-1}) coincides with the row in the submatrix $\langle \boldsymbol{B}^n, i_0(\boldsymbol{B}^{n-1})\rangle$ of G_n given by $\langle i_1(b_i), i_0(\boldsymbol{B}^{n-1})\rangle$.

Reorder the elements of $\boldsymbol{B}^n$ so that $\langle i_0(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle$ forms the upper left block of G_n and $\langle i_1(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle$ forms a block directly underneath it:

$$G_n = \begin{bmatrix} \langle i_0(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle & * & * & * & * & * \\ \langle i_1(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{bmatrix} = \begin{bmatrix} \langle i_0(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle & * & * & * & * & * \\ G_{n-1} & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{bmatrix}.$$

Lemma 5.3 implies that every row of $\langle \boldsymbol{B}^n, i_0(\boldsymbol{B}^{n-1})\rangle$ is a multiple of some row in G_{n-1}. Let $j_1, \dots, j_k$ denote the indices of all rows of $\langle \boldsymbol{B}^n, i_0(\boldsymbol{B}^{n-1})\rangle$ other than those in $\langle i_1(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle$. Let G_n' be the matrix obtained by properly subtracting multiples of rows in $\langle i_1(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle$ from rows $j_1, \dots, j_k$ of G_n

so that the submatrix obtained by restricting G'_n to rows $j_1, \ldots, j_k$ and columns corresponding to states in $i_0(\boldsymbol{B}^{n-1})$ is equal to 0:

$$G'_n = \begin{bmatrix} 0 & * & * & * & * & * \\ G_{n-1} & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & * & * & * & * & * \end{bmatrix}.$$

Thus, G'_n restricted to the columns corresponding to states in $i_0(\boldsymbol{B}^{n-1})$ contains precisely $n\binom{2n-2}{n-1}$ nonzero rows, each equal to some unique row of G_{n-1}. The determinant of this submatrix is equal to $\det G_{n-1}$. Since $\det G_{n-1}$ divides $\det G'_n$ and $\det G'_n = \det G_n$, this completes the proof. □

6. Further relations between $\det G_{n-1}$ and $\det G_n$

As noted in the previous proof, there is a submatrix of G_n equal to G_{n-1}. We will now focus on identifying multiple nonoverlapping submatrices in G_n equal to multiples of G_{n-1}. This will help in simplifying the computation of $\det G_n$. We start with a detailed justification of the first assertion in the proof of Theorem 5.4:

Lemma 6.1. *For any $b_i, b_j \in \boldsymbol{B}^{n-1}$, $\langle i_0(b_i), i_1(b_j)\rangle = \langle i_1(b_i), i_0(b_j)\rangle = \langle b_i, b_j\rangle$.*

Proof. We begin with the equality $\langle i_1(b_i), i_0(b_j)\rangle = \langle b_i, b_j\rangle$. By Proposition 5.2, $i_1(b_i) \circ i_0(b_j)^* = p_0 i_1(b_i) \circ b_j^*$, so it suffices to prove that

$$p_0 i_1(b_i) = p_0 r_{\pi/n} i_0 {r_{\pi/(n-1)}}^{-1}(b_i) = b_i.$$

This is demonstrated pictorially in Figure 13.

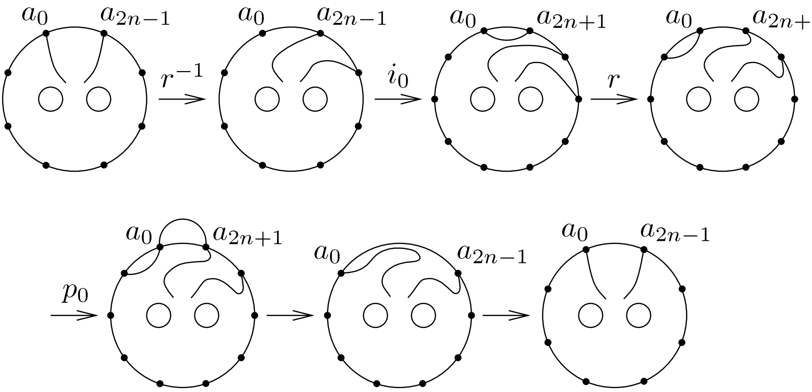


Figure 13. Proof that $p_0 \circ i_1$ is the identity.

Thus, $\langle i_1(b_i), i_0(b_j)\rangle = \langle b_i, b_j\rangle$. Recall that $\langle b_i, b_j\rangle = h_t(\langle b_j, b_i\rangle)$. From this and the previous equality, it follows that

$$\langle i_0(b_i), i_1(b_j)\rangle = h_t(\langle i_1(b_j), i_0(b_i)\rangle) = h_t(\langle b_j, b_i\rangle) = {h_t}^2(\langle b_i, b_j\rangle) = \langle b_i, b_j\rangle. \quad \square$$

Corollary 6.2. $\langle i_0(\boldsymbol{B}^{n-1}), i_1(\boldsymbol{B}^{n-1})\rangle = \langle i_1(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle = G_{n-1}$.

Lemma 6.3. *For any* $b_i, b_j \in \boldsymbol{B}^{n-1}$,

$$\langle i_0(b_i), i_0(b_j)\rangle = \langle i_1(b_i), i_1(b_j)\rangle = d\langle b_i, b_j\rangle.$$

Proof. $i_0(b_i) \circ i_0(b_j)^*$ is composed of $b_i \circ b_j^*$ in addition to a chord close to the boundary glued with its inverse. These two chords form a trivial loop. Thus, $\langle i_0(b_i), i_0(b_j)\rangle = d\langle b_i, b_j\rangle$ for all $b_i, b_j \in \boldsymbol{B}^{n-1}$.

By symmetry, $\langle i_1(\boldsymbol{B}^{n-1}), i_1(\boldsymbol{B}^{n-1})\rangle = dG_{n-1}$. □

Corollary 6.4. $\langle i_0(\boldsymbol{B}^{n-1}), i_0(\boldsymbol{B}^{n-1})\rangle = \langle i_1(\boldsymbol{B}^{n-1}), i_1(\boldsymbol{B}^{n-1})\rangle = dG_{n-1}$.

Using these facts, we can construct from G_n a $(|B_n|-2|B_{n-1}|)\times(|B_n|-2|B_{n-1}|)$ matrix whose determinant is equal to

$$\frac{\det G_n}{(1-d^2)^{n\binom{2n-2}{n-1}}}(\det G_{n-1})^2.$$

This allows us to compute $\det G_n$ with greater ease, assuming we know $\det G_{n-1}$. This process is shown in the next theorem.

Theorem 6.5. *There is a nonnegative integer*[3] k *such that*, *for all integers* $n > 1$,

$$\det {G_{n-1}}^2 \ \textit{divides}\ \det G_n(1-d^2)^k.$$

Proof. Order the elements of $\boldsymbol{B}^n$ (or equivalently, the rows and columns of G_n), as shown in Theorem 5.4. Apply the procedure from Theorem 5.4 to construct G_n', whose form is roughly

$$G_n' = \begin{bmatrix} 0 & (1-d^2)G_{n-1} & * & * & * & * \\ G_{n-1} & dG_{n-1} & * & * & * & * \\ \begin{matrix}0\\0\\0\\0\end{matrix} & \boxed{\begin{matrix}*\\*\\*\\*\end{matrix}} & \begin{matrix}*\\*\\*\\*\end{matrix} & \begin{matrix}*\\*\\*\\*\end{matrix} & \begin{matrix}*\\*\\*\\*\end{matrix} & \begin{matrix}*\\*\\*\\*\end{matrix} \end{bmatrix}.$$

Consider the block in G_n' whose columns correspond to states in $i_1(\boldsymbol{B}^{n-1})$ and whose rows correspond to states in neither $i_0(\boldsymbol{B}^{n-1})$ nor $i_1(\boldsymbol{B}^{n-1})$ (boxed above). Every row in this submatrix is a linear combination of two rows from G_{n-1}. More

[3] Clearly this integer is bounded above by $(n+1)\binom{2n}{n}$, or even better, by $|\boldsymbol{B}^n|-2|\boldsymbol{B}^{n-1}|$. Better bounds are possible, but we do not address them in this paper.

precisely, each row is of the form $a_1 l_1 - a_2 d l_2$, where l_1 and l_2 are two rows, not necessarily distinct, in G_{n-1}, and $a_1, a_2 \in \{1, d, x_1, y_1, z_2\}$. If we assume $1-d^2$ is invertible in our ring (for example, if we consider a ring of rational functions), then each row is a linear combination of two rows from $(1-d^2)G_{n-1}$. We then simplify G'_n as follows.

Let G''_n be the matrix obtained by properly subtracting linear combinations of the first $n\binom{2n-2}{n-1}$ rows of G'_n from the rows which correspond to states in neither $i_0(\boldsymbol{B}^{n-1})$ nor $i_1(\boldsymbol{B}^{n-1})$ so that the submatrix obtained by restricting G''_n to columns corresponding to states in $i_1(\boldsymbol{B}^{n-1})$ and rows corresponding to states in neither $i_0(\boldsymbol{B}^{n-1})$ nor $i_1(\boldsymbol{B}^{n-1})$ is equal to 0:

$$G''_n = \left[\begin{array}{cccccc}
0 & (1-d^2)G_{n-1} & * & * & * & * \\
G_{n-1} & dG_{n-1} & * & * & * & * \\ \cline{3-6}
0 & 0 & \multicolumn{1}{|c}{*} & * & * & \multicolumn{1}{c|}{*} \\
0 & 0 & \multicolumn{1}{|c}{*} & * & * & \multicolumn{1}{c|}{*} \\
0 & 0 & \multicolumn{1}{|c}{*} & * & * & \multicolumn{1}{c|}{*} \\
0 & 0 & \multicolumn{1}{|c}{*} & * & * & \multicolumn{1}{c|}{*} \\ \cline{3-6}
\end{array}\right].$$

The block decomposition so far proves that $\det G''_n$ equals $(1-d^2)^{n\binom{2n-2}{n-1}}(\det G_{n-1})^2$ times the determinant of the boxed block, which we denote by $\overline{G}_n$. The latter contains a power of $(1-d^2)^{-1}$, whose degree is unspecified. Thus,

$$\det {G_{n-1}}^2 \;\text{ divides }\; \det G''_n (1-d^2)^k,$$

for some integer $k \geq 0$. We remind the reader that G''_n is obtained from G'_n via determinant-preserving operations, and hence $\det G'_n = \det G_n$. □

Note that if $\det \overline{G}_n$ has fewer than $n\binom{2n-2}{n-1}$ powers of $(1-d^2)^{-1}$, then

$$\det {G_{n-1}}^2 \;\text{ divides }\; \det G_n.$$

It remains an open problem as to whether the former is true. For an example of this decomposition, we mention the equality

$$\det \overline{G}_2 = \frac{\det G_2}{(1-d^2)^4 \det {G_1}^2}.$$

7. Future directions

In this section, we discuss briefly generalizations of the Gram determinant and present a number of open questions and conjectures.

The case of a disk with k holes. We can generalize our setup by considering F^n_k, a unit disk with k holes, in addition to $2n$ points, $a_0, \ldots, a_{2n-1}$, arranged in a similar way to points in F_2^n. For $b_i, b_j \in \boldsymbol{B}_k^n$, let $b_i \circ b_j^*$ be defined in the same way as before. Each paired diagram $b_i \circ b_j^*$ consists of up to n closed curves on the 2-sphere with $2k$ holes. Let $\mathcal{S}$ denote the set of all $2k$ holes. We differentiate between the closed curves based on how they partition $\mathcal{S}$. We define a bilinear form by counting the multiplicities of each type of closed curve in the paired diagram. In the case $k=2$, we assigned to each paired diagram a corresponding element in a polynomial ring of eight variables, each variable representing a type of closed curve. In the general case, the number of types of closed curves is equal to

$$\frac{2^{|S|}}{2} = \frac{2^{2k}}{2} = 2^{2k-1},$$

so we can define the Gram matrix of the bilinear form for a disk with k holes and $2n$ points with $(n+1)^{k-1}\binom{2n}{n} \times (n+1)^{k-1}\binom{2n}{n}$ entries, each belonging to a polynomial ring of 2^{2k-1} variables. We denote this Gram matrix by $G_n^{F_k}$. For $n=1$ and $k=3$, we can easily write this 8×8 Gram matrix. For purposes of notation, let us denote the holes in $F_{0,3}^n$ by ∂_1, ∂_2 and ∂_3, and their inversions by ∂_{-1}, ∂_{-2} and ∂_{-3}, respectively. Hence, each closed curve in the surface encloses some subset of $\mathcal{S} = \{\partial_1, \partial_{-1}, \partial_2, \partial_{-2}, \partial_3, \partial_{-3}\}$. Let x_{a_1,a_2,a_3} denote a curve separating the set of holes $\{\partial_{a_1}, \partial_{a_2}, \partial_{a_3}\}$ from $\mathcal{S} - \{\partial_{a_1}, \partial_{a_2}, \partial_{a_3}\}$. We can similarly define x_{a_1,a_2} and x_{a_1}. The Gram matrix is then

$$G_1^{F_3} = \begin{bmatrix} d & x_{-3} & x_{-2} & x_{-2,-3} & x_{-1} & x_{-1,-3} & x_{-1,-2} & x_{1,2,3} \\ x_3 & x_{3,-3} & x_{-2,3} & x_{1,-1,2} & x_{-1,3} & x_{1,2,-2} & x_{1,2,-3} & x_{1,2} \\ x_2 & x_{2,-3} & x_{2,-2} & x_{1,-1,3} & x_{-1,2} & x_{1,-2,3} & x_{1,3,-3} & x_{1,3} \\ x_{2,3} & x_{1,-1,-2} & x_{1,-1,-3} & x_{1,-1} & x_{1,-2,-3} & x_{1,-2} & x_{1,-3} & x_1 \\ x_1 & x_{1,-3} & x_{1,-2} & x_{1,-2,-3} & x_{1,-1} & x_{1,-1,-3} & x_{1,-1,-2} & x_{2,3} \\ x_{1,3} & x_{1,3,-3} & x_{1,-2,3} & x_{-1,2} & x_{1,-1,3} & x_{2,-2} & x_{2,-3} & x_2 \\ x_{1,2} & x_{1,2,-3} & x_{1,2,-2} & x_{1,-3} & x_{1,-1,2} & x_{-2,3} & x_{3,-3} & x_3 \\ x_{1,2,3} & x_{-1,-2} & x_{-1,-3} & x_{-1} & x_{-2,-3} & x_{-2} & x_{-3} & d \end{bmatrix}.$$

It would be tempting to conjecture that the determinant of the matrix above has a straightforward decomposition of the form $(u+v)(u-v)$. We found that this is the case when any two variables of the form x_{a_1} and x_{a_1,a_2} are replaced by 0; explicitly, we have, with $a_1, a_2 \in \{-3, -2, -1, 1, 2, 3\}$,

$$\begin{aligned} &\det G_1^{F_3}|_{x_{a_1}=x_{a_1,a_2}=0} \\ &\quad = -(d - x_{1,2,3})(d + x_{1,2,3}) \\ &\qquad \times (x_{1,2,-2}x_{1,-1,3}x_{1,-1,-2} + x_{1,3,-3}x_{1,-1,2}x_{1,-1,-3} - x_{1,2,-3}x_{1,-1,3}x_{1,-1,-3} \\ &\qquad\qquad - x_{1,-1,-2}x_{1,-1,-2}x_{1,-2,3} - x_{1,2,-2}x_{1,3,-3}x_{1,-2,-3} + x_{1,2,-3}x_{1,-2,3}x_{1,-2,-3})^2. \end{aligned}$$

In general, however, preliminary calculations suggest that $\det G_n^{F_3}$ may be an irreducible polynomial.

Finally, we observe that many of the results we have proved for $\det G_n^{F_2}$ also hold for general $\det G_n^{F_k}$. For example, $\det G_n^{F_k}$ is nonzero and divides $\det G_n^{F_{k+1}}$. In the specific case of $\det G_n^{F_3}$, we conjecture that the diagonal term is of the form $\delta(n) = d^{\alpha(n)}(x_{1,-1}x_{2,-2}x_{3,-3})^{\beta(n)}$, where

$$\alpha(n) + 3\beta(n) = n(n+1)^2\tbinom{2n}{n} \quad \text{and} \quad \beta(n) = n(n+1)4^{n-1}.$$

***Speculation on the factorization of* $\det G_n$.** Section 5 establishes that

$$\det G_{n-1} \text{ divides } \det G_n,$$

but we conjecture that there are many more powers of $\det G_{n-1}$ in $\det G_n$. Indeed, even in the base case, $\det {G_1}^k$ divides $\det G_2$ for k up to 4. Finding the maximal power of $\det G_{n-1}$ in $\det G_n$ in the general case is an open problem and can be helpful toward computing the full decomposition of $\det G_n$.

Examining the terms of highest degree in $\det G_n$, that is, $h(\det G_n)$ may also yield helpful hints toward the full decomposition. In particular, we note that

$$\det {G_1}^4 \text{ divides } h(\det G_2) \quad \text{and} \quad \frac{h(\det G_2)^6}{\det {G_1}^9} \text{ divides } h(\det G_3).$$

We can conjecture that $(\det {G_2}^6)/(\det {G_1}^9)$ divides $\det G_3$, from which it follows that $\det {G_1}^{15}$ divides $\det G_3$. We therefore offer the following conjecture:

Conjecture 7.1. $\det {G_1}^{\binom{2n}{n-1}}$ divides $\det G_n$ for $n \geq 1$.

The next conjecture is motivated by observations of $\det G_1$ and $\det G_2$.

Conjecture 7.2. Let H_n denote the product of factors of $\det G_n$ not in $\det G_{n-1}$. Then ${H_{n-1}}^{2n}$ divides $\det G_n$.

Conjecture 7.3. Let, as before, $R = \mathbb{Z}[d, x_1, x_2, y_1, y_2, z_1, z_2, z_3]$, and let R_1 be the subgroup of R of elements invariant under h_1, h_2, h_t, and g_1, g_2, g_3. Similarly, let R_2 be the subgroup of R composed of elements $w \in R$ such that

$$h_1(w) = h_2(w) = -w \quad \text{and} \quad h_t(w) = g_1(w) = g_2(w) = g_3(w).$$

Then:

(1) $\det G_n = u^2 - v^2$, where $u \in R_1$ and $v \in R_2$.

(2) $\det G_n = \prod_\alpha (u_\alpha^2 - v_\alpha^2)$, where $u_\alpha \in R_1$ and $v_\alpha \in R_2$, and $u_\alpha - v_\alpha$ and $u_\alpha + v_\alpha$ are irreducible polynomials.

(3) $\det G_n = \prod_{i=1}^n ({u_i}^2 - {v_i}^2)^{\binom{2n}{n-i}}$, where $u_i \in R_1$ and $v_i \in R_2$.

Notice that if $w_1 = u_1^2 - v_1^2$ and $w_2 = u_2^2 - v_2^2$, then

$$w_1 w_2 = (u_1u_2 + v_1v_2)^2 - (u_1v_2 + u_2v_1)^2.$$

We have little confidence in Conjecture 7.3(3). It is closely, maybe too closely, influenced by the case of $\det G_n^{F_1}$, the Gram determinant of type B:

Theorem 7.4 [Martin and Saleur 1993; Chen and Przytycki 2009].

$$\det G_n^{F_1} = \prod_{i=1}^{n} \left(T_i(d)^2 - a^2\right)^{\binom{2n}{n-i}},$$

where $T_i(d)$ is the Chebyshev polynomial of the first kind (recursively defined by $T_0 = 2$, $T_1 = d$, $T_i = d\,T_{i-1} - T_{i-2}$), and d and a correspond to the trivial and the nontrivial curves in the annulus F_1, respectively.

References

[Birkhoff and Lewis 1946] G. D. Birkhoff and D. C. Lewis, "Chromatic polynomials", *Trans. Amer. Math. Soc.* **60** (1946), 355–451. MR 8,284f Zbl 0060.41601

[Chen and Przytycki 2008] Q. Chen and J. H. Przytycki, "The Gram matrix of a Temperley–Lieb algebra is similar to the matrix of chromatic joins", *Commun. Contemp. Math.* **10**:suppl. 1 (2008), 849–855. MR 2009m:05011 Zbl 1158.57012

[Chen and Przytycki 2009] Q. Chen and J. H. Przytycki, "The Gram determinant of the type B Temperley–Lieb algebra", *Adv. in Appl. Math.* **43**:2 (2009), 156–161. MR 2010d:57010 Zbl 1167.57004

[Deutsch and Sagan 2006] E. Deutsch and B. E. Sagan, "Congruences for Catalan and Motzkin numbers and related sequences", *J. Number Theory* **117**:1 (2006), 191–215. MR 2006k:11031 Zbl 1163.11310

[Di Francesco 1998] P. Di Francesco, "Meander determinants", *Comm. Math. Phys.* **191**:3 (1998), 543–583. MR 99e:05007 Zbl 0923.57002

[Lickorish 1991] W. B. R. Lickorish, "Invariants for 3-manifolds from the combinatorics of the Jones polynomial", *Pacific J. Math.* **149**:2 (1991), 337–347. MR 92d:57007 Zbl 0728.57011

[Lickorish 1997] W. B. R. Lickorish, *An introduction to knot theory*, Graduate Texts in Mathematics **175**, Springer, New York, 1997. MR 98f:57015 Zbl 0886.57001

[Martin and Saleur 1993] P. Martin and H. Saleur, "On an algebraic approach to higher-dimensional statistical mechanics", *Comm. Math. Phys.* **158**:1 (1993), 155–190. MR 94k:82037 Zbl 0784.05056

[Przytycki 1999] J. H. Przytycki, "Fundamentals of Kauffman bracket skein modules", *Kobe J. Math.* **16**:1 (1999), 45–66. MR 2000i:57015 Zbl 0947.57017

[Schmidt 2004] F. Schmidt, "Problems related to type-A and type-B matrices of chromatic joins", *Adv. in Appl. Math.* **32**:1-2 (2004), 380–390. MR 2004m:06007

[Simion 2000] R. Simion, "Noncrossing partitions", *Discrete Math.* **217**:1-3 (2000), 367–409. MR 2001g:05011 Zbl 0959.05009

[Stanley 1999] R. P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, 1999. MR 2000k:05026 Zbl 0945.05006

[Tutte 1991] W. T. Tutte, "On the Birkhoff–Lewis equations", *Discrete Math.* **92**:1-3 (1991), 417–425. MR 92k:05052 Zbl 0756.05059

[Tutte 1993] W. T. Tutte, "The matrix of chromatic joins", *J. Combin. Theory Ser. B* **57**:2 (1993), 269–288. MR 94a:05144 Zbl 0793.05030

[Westbury 1995] B. W. Westbury, "The representation theory of the Temperley–Lieb algebras", *Math. Z.* **219**:4 (1995), 539–565. MR 96h:20029 Zbl 0840.16008

Received: 2008-11-30 Revised: 2010-04-27 Accepted: 2010-06-03

przytyck@gwu.edu
Department of Mathematics, The George Washington University, Washington, DC 20052, United States

xzhu@fas.harvard.edu
Department of Applied Mathematics, Harvard University, Cambridge, MA 02138, United States

The cardinality of the value sets modulo n of x^2+x^{-2} and x^2+y^2

Sara Hanrahan and Mizan Khan

(Communicated by Filip Saidak)

Consider the modular circle $\mathscr{C}_{a,n}=\{(x,y):x^2+y^2\equiv a \pmod{n},\ 0\le x,y\le n-1\}$ and the modular hyperbola $\mathscr{H}_n=\{(x,y):xy\equiv 1 \pmod{n},\ 0\le x,y\le n-1\}$. We provide explicit formulas for the cardinality of the sets

$$\{a \bmod n : \mathscr{C}_{a,n}\cap\mathscr{H}_n\neq\varnothing\} \quad\text{and}\quad \{a \bmod n : \mathscr{C}_{a,n}\neq\varnothing\}.$$

Introduction

Let $\mathscr{H}_n$ denote the *modular hyperbola*

$$\{(x,y):xy=1 \pmod{n},\ 0\le x,y\le n-1\}.$$

This simply defined discrete set of points has connections to a variety of other mathematical topics including Kloosterman sums, consecutive Farey fractions, and quasirandomness. These connections have inspired a closer look at the distribution of the points of $\mathscr{H}_n$, and many questions remain open. For a discussion of recent results and open problems on modular hyperbolas, see [Shparlinski 2007].

The propensity of the points on $\mathscr{H}_n$ to collect on lines of slope ±1 was investigated in [Eichhorn et al. 2009]. In the course of that investigation, formulas for the cardinalities of the sets

$$\{(x-y) \bmod n : (x,y)\in\mathscr{H}_n\} \qquad\text{and}\qquad \{(x+y) \bmod n : (x,y)\in\mathscr{H}_n\},$$

were derived. The techniques used to determine these formulas are elementary — within the grasp of an undergraduate mathematics major who has had a course in number theory or abstract algebra.

In this article we investigate the intersection of $\mathscr{H}_n$ with the modular circles

$$\mathscr{C}_{a,n}=\{(x,y):x^2+y^2\equiv a \pmod{n},\ 0\le x,y\le n-1\},$$

MSC2000: 11A07, 11A25.

Keywords: cardinality, value sets, modular circle, modular hyperbola.

This work was done as part of Hanrahan's undergraduate honors thesis under Khan's supervision.

and in particular we determine the cardinality of the set

$$\{a \bmod n : \mathscr{C}_{a,n} \cap \mathscr{H}_n \neq \varnothing\} = \{(x^2 + y^2) \bmod n : (x, y) \in \mathscr{H}_n\}.$$

Figure 1 contrasts the modular circle $\mathscr{C}_{1,997}$ with the modular hyperbola $\mathscr{H}_{997}$. Figure 2 shows the two superimposed, and the intersection $\mathscr{C}_{1,997} \cap \mathscr{H}_{997}$.

This short note is a concise version of SH's honors thesis. It is also a natural addendum to [Eichhorn et al. 2009], as we used the formulas found there to prove our results.

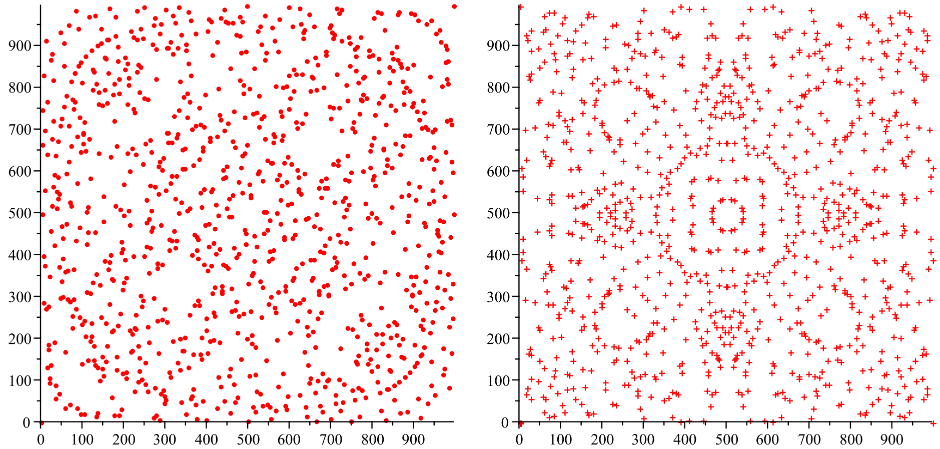


Figure 1. Left: The modular hyperbola $\mathscr{H}_{997}$. Right: The modular circle $\mathscr{C}_{1,997}$.

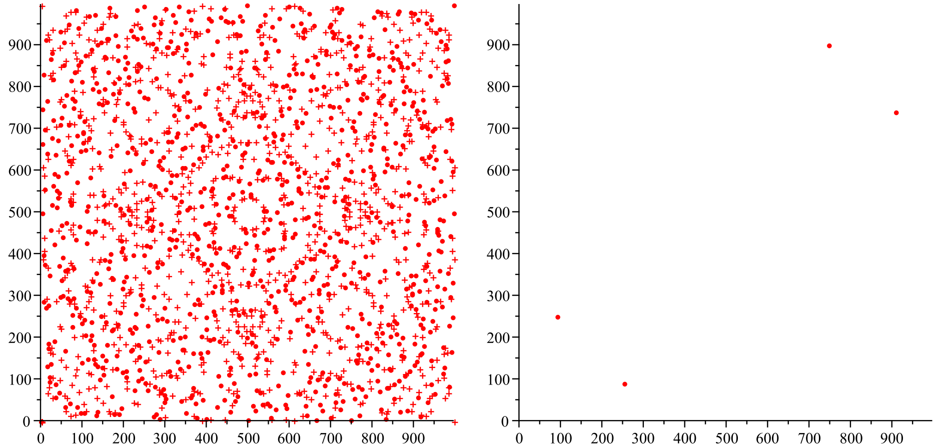


Figure 2. Left: Superposition of the preceding two sets. Points of the modular circle are represented by crosses; those of the modular hyperbola by solid circles. Right: The intersection $\mathscr{C}_{1,997} \cap \mathscr{H}_{997} = \{(91, 252), (252, 91), (745, 906), (906, 745)\}$.

1. Preliminary results

Let $f \in \mathbb{Z}[x_1, \dots, x_k]$ and let $S \subseteq \mathbb{Z}_n^k$ (where $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ is the set of integers modulo n). Then $I(f, S)$ will denote the set

$$I(f, S) = \{f(x_1, \dots, x_k) \bmod n : (x_1, \dots, x_k) \in S\}.$$

We also define two subsets of $I(f, S)$:

$$I'(f, S) = \{a : a \in I(f, S), \gcd(a, n) = 1\},$$
$$I''(f, S) = \{a : a \in I(f, S), \gcd(a, n) \neq 1\}.$$

Our first result is that the quantity $\#I(f, \mathcal{H}_n)$ is a multiplicative function of n. Furthermore, by replacing each occurrence of $\mathcal{H}_n$ with $\mathbb{Z}_n^2$ in the statement and proof of the theorem, we get that $\#I(f, \mathbb{Z}_n^2)$ is also a multiplicative function of n.

Proposition 1. *Let $f \in \mathbb{Z}[x, y]$ and define $f_n : \mathcal{H}_n \to \mathbb{Z}_n$ by*

$$f_n((x, y)) = f(x, y) \bmod n.$$

If $n = a \cdot b$ with $\gcd(a, b) = 1$, *then*

$$\#I(f, \mathcal{H}_n) = \#I(f, \mathcal{H}_a) \cdot \#I(f, \mathcal{H}_b).$$

It follows that if $n = \prod_{i=1}^m p_i^{e_i}$ is the canonical factorization of n, then

$$\#I(f, \mathcal{H}_n) = \prod_{i=1}^{m} \#I(f, \mathcal{H}_{p_i^{e_i}}). \tag{1}$$

Proof. The Chinese remainder theorem says that the map $r : \mathbb{Z}_n \to \mathbb{Z}_a \times \mathbb{Z}_b$ given by

$$r(x) = (x \bmod a, x \bmod b)$$

is an isomorphism of rings. Hence the map $R : \mathcal{H}_n \to \mathcal{H}_a \times \mathcal{H}_b$ defined by

$$R((x, y)) = ((x \bmod a, y \bmod a), (x \bmod b, y \bmod b))$$

is a bijection. The result now follows from the observation that the diagram

$$\begin{tikzcd}
\mathcal{H}_n \arrow[r, "R"] \arrow[d, "f_n"'] & \mathcal{H}_a \times \mathcal{H}_b \arrow[d, "f_a \times f_b"] \\
\mathbb{Z}_n \arrow[r, "r"] & \mathbb{Z}_a \times \mathbb{Z}_b.
\end{tikzcd}$$

commutes. □

Thus we have reduced the problem of determining formulas for $\#I(x^2+y^2, \mathcal{H}_n)$ (or $\#I(x^2+y^2, \mathbb{Z}_n^2)$) to determining them for prime powers. From this point, we shall refer to the set $I(x^2+y^2, \mathcal{H}_n)$ as $I(x^2+x^{-2}, \mathbb{Z}_n)$. All of our formulas were discovered through extensive numerical experimentation with Maple. Maple was the most valuable research tool at our disposal — only in discovering the formulas, but also in the *proving* stage. In the remainder of this section, we list the mathematical results we need to prove these formulas.

It is more convenient to work with the value set $I((x+x^{-1})^2, \mathbb{Z}_n)$ than with $I(x^2+x^{-2}, \mathbb{Z}_n)$. The following lemma justifies the change.

Lemma 2. *For any positive integer n,*

$$\#I(x^2+x^{-2}, \mathbb{Z}_n) = \#I((x+x^{-1})^2, \mathbb{Z}_n). \tag{2}$$

Proof. The map $z \mapsto (z+2) \bmod n$ defines a bijection between $I(x^2+x^{-2}, \mathbb{Z}_n)$ and $I((x+x^{-1})^2, \mathbb{Z}_n)$. □

We next state a basic criterion on the solvability of quadratic congruences modulo prime powers: $x^2 \equiv a \pmod{p^t}$.

Proposition 3 [Ireland and Rosen 1982, Propositions 4.2.3, 4.2.4, p. 46]. *Let p be prime and let a be an integer such that* $\gcd(a, p) = 1$.

(1) *Suppose $p > 2$. If the congruence $x^2 \equiv a \pmod{p}$ is solvable, then for every $t \geq 2$ the congruence $x^2 \equiv a \pmod{p^t}$ is solvable with precisely 2 distinct solutions.*

(2) *Suppose $p = 2$. If the congruence $x^2 \equiv a \pmod{2^3}$ is solvable, then for every $t \geq 3$ the congruence $x^2 \equiv a \pmod{2^t}$ is solvable with precisely 4 distinct solutions.*

Proposition 4 [Stangl 1996]. *Let p be an odd prime. Then*

$$\#I(x^2, \mathbb{Z}_{p^t}) = \frac{p^{t+1}}{2(p+1)} + (-1)^{t-1}\frac{p-1}{4(p+1)} + \frac{3}{4}. \tag{3}$$

For the special case $p = 2$ we have

$$\#I(x^2, \mathbb{Z}_{2^t}) = \frac{2^{t-1}}{3} + \frac{(-1)^{t-1}}{6} + \frac{3}{2}, \quad t \geq 2. \tag{4}$$

Proposition 5 [Eichhorn et al. 2009].

$$\#I(x+x^{-1}, \mathbb{Z}_{p^t}) = \frac{(p-3)p^{t-1}}{2} + \frac{2p^{t-1} + (-1)^{t-1}(p-1)}{2(p+1)} + \frac{3}{2}. \tag{5}$$

2. The formulas for $\#I((x+x^{-1})^2, \mathbb{Z}_{p^t})$

The central result of this paper is as follows.

Theorem 6. *For $p=2$ and $t \geq 7$,*

$$\#I((x+x^{-1})^2, \mathbb{Z}_{2^t}) = \frac{2^{t-7}}{3} + \frac{(-1)^{t-1}}{6} + \frac{3}{2}. \tag{6}$$

If $p \equiv 1 \pmod 4$ then

$$\#I((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \frac{(p-5)p^{t-1}}{4} + \frac{2p^{t-1}+(-1)^{t-1}(p-1)}{2(p+1)} + \frac{3}{2}. \tag{7}$$

If $p \equiv 3 \pmod 4$ then

$$\#I((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \frac{(p-3)p^{t-1}}{4} + \frac{2p^{t-1}+(-1)^{t-1}(p-1)}{4(p+1)} + \frac{3}{4}. \tag{8}$$

The proof occupies most of this section.

Proof of Theorem 6, case $p>2$. We will use the squaring map modulo p^t:

$$Q : I(x+x^{-1}, \mathbb{Z}_{p^t}) \to I((x+x^{-1})^2, \mathbb{Z}_{p^t}), \quad Q(z) = z^2 \bmod p^t.$$

We note that it preserves coprimeness with p:

$$Q(I'(x+x^{-1}, \mathbb{Z}_{p^t})) = I'((x+x^{-1})^2, \mathbb{Z}_{p^t}),$$
$$Q(I''(x+x^{-1}, \mathbb{Z}_{p^t})) = I''((x+x^{-1})^2, \mathbb{Z}_{p^t}).$$

Proposition 7. *Let p be an odd prime. For any $a \in I'((x+x^{-1})^2, \mathbb{Z}_{p^t})$, we have $\#Q^{-1}(\{a\}) = 2$, and consequently*

$$\#I'((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \#I'(x+x^{-1}, \mathbb{Z}_{p^t})/2. \tag{9}$$

Proof. Let a be an arbitrary element of $I'((x+x^{-1})^2, \mathbb{Z}_{p^t})$. There exists a point $(x_1, y_1) \in \mathcal{H}_{p^t}$ such that

$$(x_1+y_1)^2 \equiv a \pmod{p^t}.$$

Since $\gcd(x_1+y_1, p) = 1$,

$$x_1+y_1 \not\equiv -(x_1+y_1) \pmod{p^t};$$

hence the two distinct elements of $I'((x+x^{-1})^2, \mathbb{Z}_{p^t})$ that Q maps to a are

$$(x_1+y_1) \pmod{p^t} \quad \text{and} \quad -(x_1+y_1) \pmod{p^t}.$$

By Proposition 3, the congruence $x^2 \equiv a \pmod{p^t}$ has at most two solutions and we conclude that $\#Q^{-1}(\{a\}) = 2$. □

Proposition 8.

$$\#I''(x+x^{-1}, \mathbb{Z}_{p^t}) = \begin{cases} p^{t-1} & \text{if } p \equiv 1 \pmod 4, \\ 0 & \text{if } p \equiv 3 \pmod 4. \end{cases} \tag{10}$$

Consequently, when $p \equiv 1 \pmod 4$,

$$I''(x + x^{-1}, \mathbb{Z}_{p^t}) = \{kp : k = 0, 1, \dots, p^{t-1} - 1\}.$$

Proof. Define $s_{p^t} : \mathcal{H}_{p^t} \to \mathbb{Z}_{p^t}$ by $s_{p^t}((x, y)) = (x + y) \bmod p^t$ and let

$$\mathcal{H}''_{p^t} = \{(x, y) : (x, y) \in \mathcal{H}_{p^t} \text{ with } s_{p^t}((x, y)) \in I''(x + x^{-1}, \mathbb{Z}_{p^t})\}.$$

If $(x, y) \in \mathcal{H}''_{p^t}$, then $x + y = 0 \pmod p$ and consequently $x^2 = -1 \pmod p$. Since -1 is a quadratic residue modulo p if and only if $p \equiv 1 \pmod 4$, we obtain the second part of (10).

We now restrict our attention to primes p that are congruent to 1 modulo 4. Since $s_{p^t}(\mathcal{H}''_{p^t}) = I''(x + x^{-1}, \mathbb{Z}_{p^t})$, we prove the first part of (10) by proving the following two assertions:

(i) $\#s_{p^t}^{-1}(\{a\}) = 2$ for any $a \in I''(x + x^{-1}, \mathbb{Z}_{p^t})$.

(ii) $\#\mathcal{H}''_{p^t} = 2p^{t-1}$.

The proof of (i) is as follows. Let $(r, s) \in s_{p^t}^{-1}(\{a\})$. Then $(2r - a)$ and $(2s - a)$ are two distinct roots of the congruence

$$x^2 \equiv (a^2 - 4) \pmod{p^t}.$$

Since $p \mid a$, we have $\gcd(a^2 - 4, p) = 1$. Hence by Proposition 3

$$x^2 \equiv (a^2 - 4) \pmod{p^t}$$

cannot have more than two roots. Consequently $s_{p^t}^{-1}(\{a\}) = \{(r, s), (s, r)\}$.

We now prove (ii). Let (r, s) be an arbitrary element of $\mathcal{H}''_{p^t}$ and let

$$r = d_0 + d_1 p + d_2 p^2 + \cdots + d_{t-1} p^{t-1}$$

be the expansion of r in base p. There are only two possible choices for d_0, specifically, the two roots of $x^2 \equiv -1 \pmod p$, and for each of the other d_i's there are p possible choices: $0, 1, \dots, p - 1$. So there are $2p^{t-1}$ possible r's. Since s is completely determined by the choice of r, we conclude that $\#\mathcal{H}''_{p^t} = 2p^{t-1}$. □

Proposition 9. *If* $p \equiv 1 \pmod 4$ *then*

$$\#I''((x + x^{-1})^2, \mathbb{Z}_{p^t}) = \frac{2p^{t-1} + (-1)^{t-1}(p - 1)}{4(p + 1)} + \frac{3}{4}. \tag{11}$$

Proof. By Proposition 8

$$I''(x + x^{-1}, \mathbb{Z}_{p^t}) = \{kp : 0 \le k \le p^{t-1} - 1\}.$$

Consequently,

$$\begin{aligned} I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) &= Q(I''(x+x^{-1}, \mathbb{Z}_{p^t})) \\ &= Q(\{kp : 0 \le k \le p^{t-1}-1\}) = \{j^2 \bmod p^t : p \mid j\}. \end{aligned}$$

Therefore,

$$\#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \#\{k^2 \bmod p^t\} - \#\{k^2 \bmod p^t : \gcd(k, p) = 1\}.$$

Combining Stangl's formula (3) with the standard result that the number of quadratic residues modulo p^t is $(p^t - p^{t-1})/2$, we obtain

$$\#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \frac{2p^{t-1}+(-1)^{t-1}(p-1)}{4(p+1)} + \frac{3}{4},$$

which proves Proposition 9. □

We are now ready to prove formulas (7) and (8). We have

$$\begin{aligned} &\#I((x+x^{-1})^2, \mathbb{Z}_{p^t}) \\ &\quad = \#I'((x+x^{-1})^2, \mathbb{Z}_{p^t}) + \#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) \\ &\quad = \frac{\#I'(x+x^{-1}, \mathbb{Z}_{p^t})}{2} + \#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) \\ &\quad = \frac{\#I(x+x^{-1}, \mathbb{Z}_{p^t})}{2} - \frac{\#I''(x+x^{-1}, \mathbb{Z}_{p^t})}{2} + \#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}). \end{aligned}$$

Formula (5) is

$$\#I(x+x^{-1}, \mathbb{Z}_{p^t}) = \frac{(p-3)p^{t-1}}{2} + \frac{2p^{t-1}+(-1)^{t-1}(p-1)}{2(p+1)} + \frac{3}{2}.$$

If $p \equiv 3 \pmod 4$, then $\#I''(x+x^{-1}, \mathbb{Z}_{p^t}) = \#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) = 0$ by (10). If $p \equiv 1 \pmod 4$, then

$$\#I''(x+x^{-1}, \mathbb{Z}_{p^t}) = p^{t-1}$$

and

$$\#I''((x+x^{-1})^2, \mathbb{Z}_{p^t}) = \frac{2p^{t-1}+(-1)^{t-1}(p-1)}{4(p+1)} + \frac{3}{4},$$

by (10) and (11). We complete the proof with simple algebraic computations. □

Proof of Theorem 6, case $p = 2$. Interestingly this was the most difficult and time consuming part. It was only through experimenting with Maple that we discovered the map f (defined below) that allowed us to prove the formula for powers of 2.

Proposition 10. *Let $t \ge 3$. The image of the map*

$$f : I(x^2, \mathbb{Z}_{2^t}) \to \{0, 1, \dots, 2^{t+6}-1\}$$

given by

$$f(k^2)=(64k^2+4) \bmod 2^{t+6}$$

is $I((x+x^{-1})^2,\mathbb{Z}_{2^{t+6}})$. *Since* f *is injective we conclude that*

$$\#I((x+x^{-1})^2,\mathbb{Z}_{2^{t+6}})=\#I(x^2,\mathbb{Z}_{2^t}). \tag{12}$$

Proof. First we show that $I((x+x^{-1})^2,\mathbb{Z}_{2^{t+6}})\subseteq \operatorname{Image}(f)$. Let $(x,y)\in\mathcal{H}_{2^{t+6}}$. We can write

$$x=8x_1+a \quad\text{and}\quad y=8y_1+a,$$

with $0\le x_1, y_1<2^{t+3}$ and $a=1,3,5$ or 7. (We are using the fact that each element in $\mathbb{Z}_8^*$ is its own inverse.) The following calculation now shows that $(x+y)^2 \bmod 2^{t+6}\in\operatorname{Image}(f)$.

$$\begin{aligned}(x+y)^2&=(8x_1+8y_1+2a)^2\\&=64x_1^2-128x_1y_1+64y_1^2+256x_1y_1+32x_1a+32y_1a+4a^2\\&=64(x_1-y_1)^2+4(64x_1y_1+8x_1a+8y_1a+a^2)\\&=64(x_1-y_1)^2+4xy\\&\equiv(64(x_1-y_1)^2+4)\pmod{2^{t+6}}.\end{aligned}$$

To show the reverse inclusion, let $k^2\in I(x^2,\mathbb{Z}_{2^t})$. By Proposition 3 the congruence

$$x^2\equiv 16k^2+1 \pmod{2^n}$$

has a solution for all values of n. Let l be any integer such that $l^2=16k^2+1 \pmod{2^{t+6}}$, and let

$$x=(l-4k) \bmod 2^{t+6},\qquad y=(l+4k) \bmod 2^{t+6}.$$

The immediate observations that $(x,y)\in\mathcal{H}_{2^{t+6}}$ and

$$(x+y)^2\equiv 4l^2\equiv 64k^2+4 \pmod{2^{t+6}}$$

complete the proof. □

Now the formula (6) for $\#I((x+x^{-1})^2,\mathbb{Z}_{2^t})$ is obtained by combining (2), (12) and (16). This concludes the proof of Theorem 6. □

We can also derive the formula for $\#I(x^2+x^{-2},\mathbb{Z}_p)$ as a special case of an old formula for pairs of quadratic residues.

Theorem 11 [Berndt et al. 1998, Theorem 6.3.1, page 197]. *Let p be an odd prime and let c be an integer relatively prime to p. Let $\epsilon_1=\pm1$ and $\epsilon_2=\pm1$. Then*

$$\#\Big\{n:0\le n<p,\ \Big(\frac{n}{p}\Big)=\epsilon_1,\ \Big(\frac{n+c}{p}\Big)=\epsilon_2\Big\}\\ =\frac{1}{4}\Big\{p-2\epsilon_1\Big(\frac{-c}{p}\Big)-\epsilon_2\Big(\frac{c}{p}\Big)-\epsilon_1\epsilon_2\Big\}. \tag{13}$$

The special case of this formula with $\epsilon_1 = \epsilon_2 = c = 1$ was first published by Aladov in 1896. The connection between (13) and $\#I(x^2+x^{-2}, \mathbb{Z}_p)$ is as follows.

Theorem 12. *Let $a \in \mathbb{Z}$ with $\gcd(a^2-4, n) = 1$. Then $\mathscr{C}_{a,n} \cap \mathscr{H}_n \neq \varnothing$ if and only if for every prime, p, in the canonical factorization of n we have*

$$\left(\frac{a-2}{p}\right) = \left(\frac{a+2}{p}\right) = 1. \tag{14}$$

Consequently,

$$\#I(x^2+x^{-2}, \mathbb{Z}_p) = \#\left\{a \,:\, 0 \le a < p,\ \left(\frac{a-2}{p}\right) = \left(\frac{a+2}{p}\right) = 1\right\} + 1.$$

Proof. For the "only if" part, let $(r, s) \in \mathscr{C}_{a,n} \cap \mathscr{H}_n$ and let p be an arbitrary prime divisor of n. So, $(r-s)^2 \equiv a-2 \pmod{p}$ and $(r+s)^2 \equiv a+2 \pmod{p}$, which leads immediately to (14).

To prove the converse, let $n = \prod_{i=1}^{t} p_i^{e_i}$ be the canonical factorization of n. By Proposition 3, we can lift the square roots (modulo p) of $(a-2)$ and $(a+2)$ to the e_ith power, $p_i^{e_i}$. Let $s_i = \sqrt{a-2} \pmod{p_i^{e_i}}$, and $r_i = \sqrt{a+2} \pmod{p_i^{e_i}}$. Then

$$2^{-1} \cdot (r_i + s_i, r_i - s_i) \in \mathscr{C}_{p_i^{e_i}} \cap \mathscr{H}_{p_i^{e_i}},$$

where 2^{-1} denotes the inverse of 2 modulo $p_i^{e_i}$. Now invoke the Chinese remainder theorem to determine integers r and s such that

$$r \equiv r_i \pmod{p_i^{e_i}} \text{ and } s \equiv s_i \pmod{p_i^{e_i}} \qquad \text{for } i = 1, \ldots, t.$$

Clearly $(r, s) \in \mathscr{C}_n \cap \mathscr{H}_n$. □

3. The formulas for $\#I(x^2+y^2, \mathbb{Z}_{p^t}^2)$

We now determine the formulas for $\#I(x^2+y^2, \mathbb{Z}_{p^t}^2)$ to contrast them to

$$\#I(x^2+x^{-2}, \mathbb{Z}_{p^t}).$$

Theorem 13. *Let p be an odd prime. Then*

$$\#I(x^2+y^2, \mathbb{Z}_{p^t}^2) = \begin{cases} p^t & \text{if } p \equiv 1 \pmod{4}, \\ p & \text{if } p \equiv 3 \pmod{4} \text{ and } t = 1, \\ p^t - \sum_{j=0}^{[t/2]-1} \varphi(p^{t-1-2j}) & \text{if } p \equiv 3 \pmod{4} \text{ and } t > 1, \end{cases} \tag{15}$$

When $p = 2$ we have

$$\#I(x^2+y^2, \mathbb{Z}_{2^t}^2) = \varphi(2^t) + 1. \tag{16}$$

As is typically the case, the formula for powers of two, 2^t, will require a separate argument. We first prove (15).

Proof of formula (15). We treat each case separately.

- $p \equiv 1 \pmod 4$. Let $a \in \{0, 1, \dots, p^t - 1\}$. The simultaneous congruences

$$x - y \equiv 1 \pmod{p^t} \quad \text{and} \quad x + y \equiv a \pmod{p^t}$$

have the solutions

$$x = ((a+1) \cdot (2^{-1} \bmod p^t)) \bmod p^t,$$
$$y = ((a-1) \cdot (2^{-1} \bmod p^t)) \bmod p^t.$$

It immediately follows that $x^2 + (i_{p^t} y)^2 \equiv a \pmod{p^t}$, where

$$i_{p^t}^2 \equiv -1 \pmod{p^t}.$$

- $p \equiv 3 \pmod 4$, $t = 1$. Let $a \in \{0, 1, \dots, p-1\}$. By (3), $\# I(x^2, \mathbb{Z}_p) = (p+1)/2$ and therefore $\#(a - I(x^2, \mathbb{Z}_p)) = (p+1)/2$. Since

$$\# I(x^2, \mathbb{Z}_p) + \#(a - I(x^2, \mathbb{Z}_p)) = p + 1,$$

it follows that there is an element $(a - x_1^2) \in (a - I(x^2, \mathbb{Z}_p))$ and an element $x_2^2 \in I(x^2, \mathbb{Z}_p)$ such that $(a - x_1^2) \equiv x_2^2 \pmod p$.

- $p \equiv 3 \pmod 4$, $t \geq 2$. The key is to prove that an element $a \in \{0, 1, 2, \dots, p^t - 1\}$ satisfies $a \equiv x^2 + y^2 \pmod{p^t}$ if and only if $a = p^k b$, with $\gcd(p, b) = 1$ and k *even*.

($\Leftarrow$) Since p^k is a square in $\mathbb{Z}$, it is sufficient to prove this for integers a that are relatively prime to p. We argue by induction. The previous case shows that the result holds for $t = 1$. Let us assume it is true for t. So

$$a \equiv (x^2 + y^2) \pmod{p^t}.$$

If $p^{t+1} \mid (a - x^2 - y^2)$, there is nothing to prove. So let us assume that $(a - x^2 - y^2) = p^t l$, with $\gcd(l, p) = 1$. Since $\gcd(a, p) = 1$ either $\gcd(x, p) = 1$ or $\gcd(y, p) = 1$. Without loss of generality we assume the former. We now define $s \in \mathbb{Z}$, with $1 \leq s < p$, to be the solution of the congruence

$$2xs \equiv l \pmod p.$$

An immediate calculation shows that

$$a \equiv (x + sp^t)^2 + y^2 \pmod{p^{t+1}}.$$

($\Rightarrow$) We argue by contradiction. Suppose $a = p^k b$, with $a < p^t$, $\gcd(b, p) = 1$,and k odd, be the sum of two squares modulo p^t. So there are integers $x = p^{e_1} x_1$, $y = p^{e_2} y_1$, with $\gcd(x_1 y_1, p) = 1$, such that

$$p^k b \equiv (x^2 + y^2) \pmod{p^t},$$

that is,

$$p^k b \equiv (p^{2e_1}x_1^2 + p^{2e_2}y_1^2) \pmod{p^t}.$$

Since $b \not\equiv 0 \bmod p$ and k is odd we have $\min\{2e_1, 2e_2\} < k$. Without loss of generality we may assume that $e_1 \leq e_2$. We can reduce the congruence

$$p^k b \equiv (x^2 + y^2) \pmod{p^t}$$

to $p^{k-2e_1}b \equiv x_1^2 + p^{2(e_2-e_1)}y_1^2 \pmod{p^{k-2e_1}}$, which in turns reduces to

$$x_1^2 + p^{2(e_2-e_2)}y_1^2 \equiv 0 \pmod{p}.$$

Since $x_1 \not\equiv 0 \pmod{p}$ we must have $p^{2(e_2-e_2)}y_1^2 \not\equiv 0 \pmod{p}$, that is $e_2 = e_1$, and consequently $(x_1^2 + y_1^2) \equiv 0 \pmod{p}$, with $\gcd(x_1y_1, p) = 1$. But this gives us the contradiction that $x^2 \equiv -1 \pmod{p}$ is solvable for a prime p with $p \equiv 3 \pmod{4}$. This concludes the proof of (15). □

Proposition 14. *Let $t \geq 3$ and $0 < m < 2^t$. Then $m \in I(x^2+y^2, \mathbb{Z}_{2^t}^2)$ if and only if $m = 2^j \cdot a$, with $j < t$ and $a \equiv 1 \pmod{4}$.*

Proof. ($\Leftarrow$) Let $a \equiv 1 \pmod{4}$. Since 2^j is a sum of squares (in $\mathbb{Z}$) we only need to show that a is a sum of two squares modulo 2^t. If $a \equiv 1 \pmod{8}$ then a is a square modulo 2^t by Proposition 3. If $a \equiv 5 \pmod{8}$, then $a-4 \equiv 1 \pmod{8}$ and is therefore a square modulo 2^t. Consequently a is a sum of two squares modulo 2^t.

($\Rightarrow$) We now assume that $a \equiv 3 \pmod{4}$ and argue by contradiction. Let

$$x^2 + y^2 \equiv m \pmod{2^t}.$$

We look at four possible cases.

(1) $j = 0$: We obtain the contradiction that

$$x^2 + y^2 \equiv 3 \pmod{4}.$$

(2) $j = 1$: We obtain the contradiction that

$$x^2 + y^2 \equiv 6 \pmod{8}.$$

(3) $j \geq 2$, $j \leq (t-2)$: We have $x = 2^{e_1} \cdot x_1$ and $y = 2^{e_2} \cdot y_1$, with x_1, y_1 odd and $j = \min\{2e_1, 2e_2\}$. Without loss of generality we may assume that $e_1 \leq e_2$. We now obtain the contradiction

$$x_1^2 + 4^{e_2-e_1}y_1^2 \equiv a \equiv 3 \pmod{4}.$$

(4) $j = t-1$: Then

$$m = 2^{t-1} \cdot a \geq 2^{t-1} \cdot 3 > 2^t,$$

contradicting the fact that the elements of $I(x^2+y^2, \mathbb{Z}_{2^t}^2)$ are less than 2^t. □

Proof of formula (16). Let M_t denote the set

$$M_t = \{m : 0 < m < 2^t,\ m = 2^j \cdot a,\ j < t,\ a \equiv 1 \pmod 4\}.$$

In our previous proposition we proved that

$$I(x^2 + y^2, \mathbb{Z}^2_{2^t}) \setminus \{0\} = M_t.$$

We now make the following two observations about elements in M_t:

(i) If $m \in M_t$, then $(m + 2^t) \in M_{t+1}$ provided $m \neq 2^{t-1}$.

(ii) If $m \in M_{t+1}$ with $m > 2^t$, then $(m - 2^t) \in M_t$.

From these two observations we conclude that

$$M_{t+1} \setminus \{2^t\} = M_t \cup \{m + 2^t : m \in M_t \setminus \{2^{t-1}\}\},$$

and consequently $\#M_{t+1} = 2 \cdot \#M_t$. An inductive argument now proves that $\#M_t = \varphi(2^t)$ and therefore $\#I(x^2 + y^2, \mathbb{Z}^2_{2^t}) = \varphi(2^t) + 1$. □

References

[Berndt et al. 1998] B. C. Berndt, R. J. Evans, and K. S. Williams, *Gauss and Jacobi sums*, CMS Monographs and Advanced Texts **21**, Wiley, New York, 1998. MR 99d:11092 Zbl 0906.11001

[Eichhorn et al. 2009] D. Eichhorn, M. R. Khan, A. H. Stein, and C. L. Yankov, "Sums and differences of the coordinates of points on modular hyperbolas", pp. 17–39 in *Combinatorial number theory* (Carrollton, GA, 2007), edited by B. Landman et al., de Gruyter, Berlin, 2009. Also published in *Integers: Electronic J. Combin. Number Theory*, **9** supplement (2009), article 3. MR 2010i:11149 Zbl 1178.11004

[Ireland and Rosen 1982] K. F. Ireland and M. I. Rosen, *A classical introduction to modern number theory*, Grad. Texts in Math. **84**, Springer, New York, 1982. MR 83g:12001 Zbl 0482.10001

[Shparlinski 2007] I. E. Shparlinski, "Distribution of points on modular hyperbolas", pp. 155–189 in *Sailing on the sea of number theory: Proc. 4th China-Japan Seminar on Number Theory* (Weihai, 2006), edited by S. Kanemitsu and J.-Y. Liu, Ser. Number Theory Appl. **2**, World Sci. Publ., Hackensack, NJ, 2007. MR 2008m:11162 Zbl 1175.11044

[Stangl 1996] W. D. Stangl, "Counting squares in Z_n", *Math. Mag.* **69**:4 (1996), 285–289. MR 1424442 Zbl 1055.11500

Received: 2009-02-08 Accepted: 2010-06-21

hanrahans@stu.easternct.edu
Department of Mathematics and Computer Science,
Eastern Connecticut State University,
Willimantic, CT 06226, United States

khanm@easternct.edu
Department of Mathematics and Computer Science,
Eastern Connecticut State University,
Willimantic, CT 06226, United States

Minimal k-rankings for prism graphs

Juan Ortiz, Andrew Zemke, Hala King, Darren Narayan and Mirko Horňák

(Communicated by Vadim Ponomarenko)

We determine rank numbers for the prism graph $P_2 \times C_n$ (P_2 being the connected two-node graph and C_n a cycle of length n) and for the square of an even cycle.

1. Introduction

A k-ranking of a graph is a vertex labeling using integers between 1 and k inclusive such that any path between two vertices of the same rank contains a vertex of strictly larger rank. When the value of k is unimportant, we will refer to a k-ranking simply as a ranking. A ranking f is minimal if the reduction of any label violates the ranking property [Ghoshal et al. 1996]. Another definition of a minimal ranking is obtained by replacing the reduction of a label by the reduction of labels for any nonempty set of vertices. It was shown in [Jamison 2003] and [Isaak et al. 2009] that these two definitions of minimal rankings are equivalent. The *rank number* of a graph G, denoted $\chi_r(G)$ is the smallest k such that G has a minimal k-ranking.

Recall that a vertex coloring of a graph is a vertex labeling in which no two adjacent vertices have the same label. Hence a k-ranking is a restricted vertex coloring. Then the rank number is similar to the chromatic number. The *arank number of a graph* G, denoted $\psi_r(G)$, is the largest k such that G has a minimal k-ranking.

The study of the rank number was motivated by applications including the design of very large scale integration (VLSI) layout and Cholesky factorizations associated with parallel processing [de la Torre et al. 1992; Ghoshal et al. 1996; 1999;

MSC2000: primary 05C15, 05C78; secondary 05C69, 05C35.

Keywords: k-rankings, vertex coloring, prism graphs.

This research started on a National Science Foundation Research Experience for Undergraduates Program at the Rochester Institute of Technology (RIT) during the summer of 2007, which was cofunded by Department of Defense. The project was continued as part of a capstone project for Juan Ortiz at California Lutheran University under the direction of Hala King, and as an undergraduate research project at RIT by Andrew Zemke under the direction of Darren Narayan. The work of Mirko Horňák was supported by the Science and Technology Assistance Agency under contract no. APVV-0007-07, by grant VEGA 1/0428/10 and by the Agency of the Slovak Ministry of Education for the Structural Funds of the EU under the project ITMS:26220120007.

Leiserson 1980; Laskar and Pillone 2001; 2000; Sen et al. 1992]. Numerous related papers have since followed [Bodlaender et al. 1998; Hsieh 2002; Jamison 2003; Dereniowski 2006; 2004; Dereniowski and Nadolski 2006; Kostyuk and Narayan $\geq$ 2010; Kostyuk et al. 2006; Isaak et al. 2009; Novotny et al. 2009a]. Ghoshal, Laskar, and Pillone were the first to investigate minimal k-rankings [Ghoshal et al. 1999; 1996; Laskar and Pillone 2001; 2000]. The determination of the rank number and the arank number was shown to be NP-complete [Laskar and Pillone 2000]. The rank number was explored in [Bodlaender et al. 1998] where the authors showed that $\chi_r(P_n) = \lfloor \log_2 n \rfloor + 1$. Rank numbers are known for a few other graph families such as cycles, wheels, complete bipartite graphs, and split graphs [Ghoshal et al. 1996; Dereniowski 2004]. The rank number for ladder graphs $P_2 \times P_n$ and the square of a path P_n^2 were determined in [Novotny et al. 2009b].

Throughout the paper P_n will denote the path on n vertices. We use $G \times H$ to denote the *Cartesian product* of G and H. The k-th power of a path, P_n^k, has vertices $v_1, v_2, \dots, v_n$ and edges (v_i, v_j) for all i, j satisfying $|i - j| \leq k$. The k-th power of a cycle, C_n^k, is defined similarly.

In this paper we determine rank numbers for the prism graph $P_2 \times C_n$ and the square of an even cycle.

We begin by restating two elementary results from [Ghoshal et al. 1996].

Lemma 1. *In any minimal ranking of a connected graph G the highest label must be unique.*

Proof. Suppose there exist two vertices u and v that both have the highest label k. Then any path between u and v will not contain a vertex with a higher label. This is a contradiction. □

The following lemma gives a monotonicity result involving the rank number.

Lemma 2. *Let H be a subgraph of a graph G. Then $\chi_r(H) \leq \chi_r(G)$.*

Proof. The proof is straightforward. Suppose $\chi_r(H) > \chi_r(G)$. Then we could relabel the vertices of H using the corresponding labels used in the ranking of G. This produces a ranking with fewer labels, and hence a contradiction. □

1.1. *The ladder graph L_n.*

We next describe a family of graphs built using the *Cartesian product.*

Definition 3. The *Cartesian product* of G and H written $G \times H$ is the graph with vertex set $V(G) \times V(H)$ specified by putting $\{u, v\}$ adjacent to (u', v') if and only if $u = u'$ and $(v, v') \in E(H)$ or $v = v'$ and $(u, u') \in E(G)$.

An example is the ladder graph $L_n = P_2 \times P_n$, shown in Figure 1.

In this paper we investigate the family of prism graphs $P_2 \times C_n$. We will start with a ladder $P_2 \times P_n$ with n even, and insert either a $P_2 \times P_1$ or $P_2 \times P_2$ and

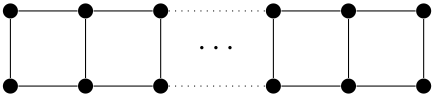


Figure 1. The ladder graph $L_n = P_2 \times P_n$.

"wrap" the ends to form a prism graph $P_2 \times C_{n+1}$ or $P_2 \times C_{n+2}$. In order for this construction to work, it is essential that in the labeling of the vertices labeled 1 of the ladder satisfies an "alternating 1's property": for each vertex v, either v is labeled 1 or all of its neighbors are labeled 1 (Figure 2). That is, the vertices labeled 1 form a particular dominating set of the graph. It was shown in [Novotny et al. 2009b] that in a minimal ranking of a ladder the 1's can be made to alternate.

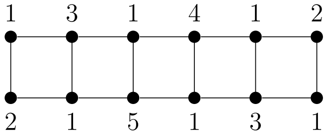


Figure 2. A graph with the alternating 1s property.

We can insert in $P_2 \times P_n$ either a 1-bridge (Figure 3, left) or a 2-bridge (Figure 3, right). In general, the bridges will contain the labels k and $k+1$ where $k-1$ is the rank of the original ladder. Our example shows the extension where $k = 6$.

In each case we insert four edges to connect the bridge to each end of the ladder. When n is even the wrapping of the ladder L_n creates a prism graph where the 1's alternate. When n is odd the 1's alternate except in one place where there are two vertices labeled 1 that are distance 3 apart (Figure 4).

Novotny et al. [2009b] determined the rank number of a ladder graph. This result is stated in our next lemma.

Lemma 4. $\chi_r(L_n) = \lfloor \log_2(n+1) \rfloor + \lfloor \log_2(n+1-2^{\lfloor \log_2 n \rfloor - 1}) \rfloor + 1$ *for* $n \geq 1$.

Applying our construction immediately gives an upper bound for the rank number of the prism graph $P_2 \times C_n$, as stated in our next theorem.

Theorem 5. *For* $k \geq 2$, *both* $\chi_r(P_2 \times C_{2k-1})$ *and* $\chi_r(P_2 \times C_{2k})$ *are bounded from above by* $r(2k-2) + 2$.

We will show later that this bound is tight.

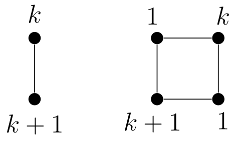


Figure 3. A 1-bridge (left) and 2-bridge (right).

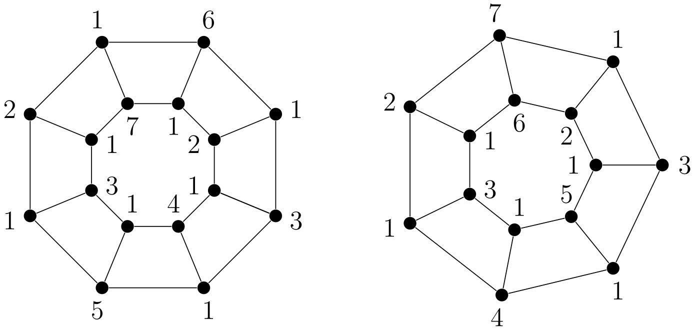


Figure 4. Prism graphs for n even (left) and n odd (even).

2. Main results

Theorem 6. *Let $l = \chi_r(P_2 \times C_n)$ where $n \geq 3$. If f is a minimal l-ranking of $P_2 \times C_n$, then $l \geq 5$ and the largest four labels of f appear exactly once.*

Proof. In the minimal ranking $f : V(P_2 \times C_n) \to \{1, 2, \dots, l\}$ every label appears at least once. Since $G = P_2 \times C_n$ is (vertex) 3-connected, any two distinct vertices of G are joined by three internally vertex disjoint paths. Hence each of the largest three labels appears exactly once in f.

Assume that $l - 3$ appears at least twice with $f(x) = f(y) = l - 3$, where $x \neq y$. We have $l \geq 5$ because the independence number of G is $2\lfloor n/2 \rfloor$ and $2\lfloor n/2 \rfloor + 3 < 2n = |V(G)|$.

Let S be a minimum-sized x, y vertex separating set. It is clear that $|V(S)| = 3$. It is well known that every 3-element separating set $\tilde{S}$ is a prism graph P is a neighborhood of a single vertex $\tilde{z} \in V(P)$ and the nontrivial component of $P - \tilde{S}$ is induced by $V(P) - (\tilde{S} \cup \{\tilde{z}\})$. Thus, there exists $z \in \{x, y\}$ such that S is the neighborhood of z. However if z has its neighbors labeled $l-2$, $l-1$, and l, then $f(x)$ can be reduced to 1, contradicting the minimality of f. □

For a positive integer n let

$$r(n) = \lfloor \log_2(n+1) \rfloor + \lfloor \log_2(n + 1 - (2^{\lfloor \log_2 n \rfloor - 1})) \rfloor + 1. \tag{1}$$

Then Lemma 4 states that $\chi_r(L_n) = \chi_r(P_2 \times P_n) = r(n)$ for $n \geq 1$.

Theorem 7. *For $k \geq 2$, we have*

$$\chi_r(P_2 \times C_{2k-1}) = \chi_r(P_2 \times C_{2k}) = \chi_r(P_2 \times P_{2k-2}) + 2 = r(2k-2) + 2.$$

Proof. By Theorem 5, both $\chi_r(P_2 \times C_{2k-1})$ and $\chi_r(P_2 \times C_{2k})$ are bounded from above by $r(2k-2) + 2$. In other words, if $m = 2k - 1$ or $2k$, then

$$\chi_r(P_2 \times C_m) \leq \chi_r(P_2 \times P_{2\lceil m/2 \rceil - 2}) + 2 = r(2\lceil m/2 \rceil - 2) + 2.$$

To prove the theorem we will show that this last inequality is in fact equality. If $k = 2$ and $m = 2k - 1$ or $2k$, then $r(2\lceil m/2\rceil - 2) + 2 = 5$. So by Theorem 6, $\chi_r(P_2 \times C_m) = 5$.

Now assume that $m = 2k - 1$ or $2k$, $k \geq 3$, and

$$\chi_r(P_2 \times C_m) = l \leq r\Big(2\Big\lceil \frac{m}{2} \Big\rceil - 2\Big) + 1. \tag{2}$$

Let f be an l-minimal ranking of $G = P_2 \times C_m$. If $k = 3$, then $5 \leq l = r(4) + 1 \leq 5$, $l = 5$, and by Theorem 6, the label 1 appears $2m - 4$ times in f. However the independence number of G equals $2\lfloor m/2\rfloor \leq m < 2m - 4$, which is a contradiction.

Let $k \geq 4$. This implies $m \geq 7$. Let i be the maximum label used at least twice. Since $r(2\lceil m/2\rceil - 2) + 1 \leq r(m-1) + 1 < 2m = |V(G)|$, such a label does exist, and $i \leq l - 4$ by Theorem 6. Consider vertices $x_1, x_2 \in V(G)$ with $f(x_1) = i = f(x_2)$, and let y_j be the neighbor of x_j that is not on the "ring" containing x_j. We will refer to this vertex as the special neighbor of x_j for $j = 1, 2$. There are two distinct subgraphs G_1, G_2 of G that are ladders with corners x_1, x_2, y_1, y_2. The restriction $f|_{V(G_j)}$ is a ranking of G_j; hence there is a minimal separating set $S_j \subseteq V(G_j)$ such that $\min f(S_j) > i$ and x_1, x_2 are in distinct components of $G_j - S_j$, $j = 1, 2$. It is easy to see that any minimal separating set that separates two "distant" corners of a ladder on at least six vertices has two vertices and is of one of the two types shown in Figure 3 (consisting of the vertices labeled k and $k + 1$). As all labels in $\{i + 1, \dots, l\}$ are used by f exactly once, any permutation of those labels yields a ranking of G. Therefore, we may suppose without loss of generality that $f(S_1) \cup f(S_2) = \{l - 3, l - 2, l - 1, l\}$. Further, let $\bar{S}_j$ be the set consisting of the vertices of S_j together with their special neighbors (so that $|\bar{S}_j|$ is 2 or 4). The graph $G - (\bar{S}_1 \cup \bar{S}_2)$ is a union of two vertex disjoint ladders H_1 and H_2. Clearly if $|V(H_1)| \geq |V(H_2)|$, then $H_1 = P_2 \times P_q$, where $q \geq \lceil (m-4)/2\rceil$. Now $f|_{V(H_1)}$ uses only labels from the set $\{1, \dots, l - 4\}$; hence, by (2),

$$\chi_r(H_1) \leq l - 4 \leq r\left(2\left\lceil \frac{m}{2}\right\rceil - 2\right) - 3. \tag{3}$$

On the other hand if s, t are positive integers with $s \leq t$, then $P_2 \times P_s$ is a subgraph of $P_2 \times P_t$. Then by Lemma 2 we have $r(s) = \chi_r(P_2 \times P_s) \leq \chi_r(P_2 \times P_t) = r(t)$. Consequently,

$$\chi_r(H_1) = \chi_r(P_2 \times P_q) = r(q) \geq r\left(\left\lceil \frac{m-4}{2}\right\rceil\right). \tag{4}$$

If m is even, then it follows from Equations (3) and (4) that

$$r(m-2) = r\Big(2 \cdot \frac{m-4}{2} + 2\Big) \geq r\Big(\frac{m-4}{2}\Big) + 3.$$

If m is odd we have

$$r(m-1)=r\left(2\cdot\frac{m-3}{2}+2\right)\geq r\left(\frac{m-3}{2}\right)+3.$$

However both cases lead to a contradiction. From (1) it is easy to see that

$$r(2n+2)-r(n)=2$$

for any positive integer n. □

Since $r(2k-3)=r(2k-2)$ for $n\geq 3$, we obtain from Theorem 7:

Theorem 8. *$\chi_r(P_2\times C_n)=\chi_r(L_{n-2})+2$ for $n\geq 4$.*

3. Rankings for other classes of graphs

We now show that the rank number of a prism graph can be used to give the rank number of the square of an even cycle. We recall some earlier facts:

Definition 9 [Ghoshal et al. 1996]. For a graph G and a set $S\subseteq V(G)$ the *reduction* of G, denoted by $G_S^\flat$, is a subgraph of G induced by $V-S$ with an edge uv in $E(G_S^\flat)$ if and only if there exists a $u-v$ path in G with all internal vertices belonging to S.

Lemma 10 [Ghoshal et al. 1996]. *Let G be a graph and let f be a minimal k-ranking of G. If*

$$S_1=\{x\in V(G):f(x)=1\}\quad\textit{and}\quad f^\flat:V(G_{S_1}^\flat)\to\{1,\dots,k-1\}$$

is defined by $f^\flat(x)=f(x)-1$, then $f^\flat$ is a minimal $(k-1)$-ranking of $G_{S_1}^\flat$.

3.1. ***The square of a cycle.*** Next we reduce even prism graphs to squares of cycles.

Theorem 11. *$\chi_r(C_n^2)=\chi_r(P_2\times C_n)$ for even $n\geq 4$.*

Proof. (Illustrated in Figure 5.) If $n=2$, the result follows from Theorem 7 which states that $\chi_r(P_2\times C_4)=5$ and from the fact that $\chi_r(C_4^2)=\chi_r(K_4)=4$.

Henceforth suppose that $n\geq 3$. Let $k=\chi_r(P_2\times C_{2n})$ and let $l=\chi_r(C_{2n}^2)$. Let f be a k-ranking of $P_2\times C_{2n}$ in which the 1's alternate. It is straightforward to see that then $(P_2\times C_{2n})_{S_1}^\flat$ is isomorphic to C_{2n}^2. Therefore, by Lemma 10 $\chi_r(C_{2n}^2)\leq k-1$.

Now let g be an l-ranking of C_{2n}^2. One can easily see that C_{2n}^2 is isomorphic to an n-sided antiprism A_n. Pick a new vertex inside each of the $2n$ triangles of A_n, join it to all three vertices of "its" triangle and delete all edges of A_n. The result is a graph G_n that is isomorphic to $P_2\times C_{2n}$. Consider the mapping $\tilde{g}:V(G_n)\to\{1,\dots,l+1\}$ defined as follows: $\tilde{g}(x)=g(x)+1$ if $x\in V(C_{2n}^2)$ and $\tilde{g}(x)=1$ if $x\in V(G_n)-V(C_{2n}^2)$. Since $\tilde{g}$ is a ranking of G_n (a simple exercise left to the reader), we have $\chi_r(P_2\times C_{2n})\leq l+1$.

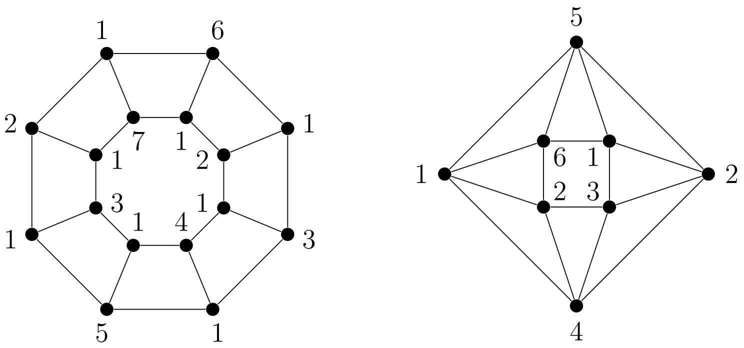


Figure 5. A minimal 7-ranking of $P_2 \times C_8$ (left) and a minimal 6-ranking of A_4 (right).

Thus $l = \chi_r(C_{2n}^2) \leq k - 1 = \chi_r(P_2 \times C_{2n}) - 1 \leq (l+1) - 1 = l$, and since both inequalities turn into equalities, we are done. □

Combining Theorems 8 and 11 gives:

Corollary 12. *Let $n \geq 4$ be even. Then*

$$\chi_r(C_n^2) = \chi_r(P_2 \times C_n) - 1 = \lfloor \log_2(n-1) \rfloor + \left\lfloor \log_2\left(n - 1 - (2^{\lfloor \log_2(n-2) \rfloor - 1})\right) \right\rfloor + 2.$$

4. Conclusion

We conclude by posing some problems for future research. In this paper we determined the rank number of $P_2 \times C_n$ using known results for the rank number of $P_2 \times P_n$. It would be interesting to determine the rank numbers for grid graphs $P_m \times P_n$ and cylinders $P_m \times C_n$. We found out recently that [Alpert $\geq$ 2010] gives rank numbers for $P_3 \times P_n$, among other results including an alternate proof of our Theorem 7.

References

[Alpert $\geq$ 2010] H. Alpert, "Rank numbers of grid graphs", *Discrete Math.* To appear.

[Bodlaender et al. 1998] H. L. Bodlaender, J. S. Deogun, K. Jansen, T. Kloks, D. Kratsch, H. Müller, and Z. Tuza, "Rankings of graphs", *SIAM J. Discrete Math.* **11**:1 (1998), 168–181. MR 99b:68143 Zbl 0907.68137

[Dereniowski 2004] D. Dereniowski, "Rank coloring of graphs", pp. 79–93 in *Graph colorings*, edited by M. Kubale, Contemp. Math. **352**, Amer. Math. Soc., Providence, RI, 2004. MR 2076991

[Dereniowski 2006] D. Dereniowski, *Parallel scheduling by graph ranking*, Ph.D. thesis, Gdańsk University of Technology, 2006.

[Dereniowski and Nadolski 2006] D. Dereniowski and A. Nadolski, "Vertex rankings of chordal graphs and weighted trees", *Inform. Process. Lett.* **98**:3 (2006), 96–100. MR 2006j:68032 Zbl 1187.68340

[Ghoshal et al. 1996] J. Ghoshal, R. Laskar, and D. Pillone, "Minimal rankings", *Networks* **28**:1 (1996), 45–53. MR 97e:05110 Zbl 0863.05071

[Ghoshal et al. 1999] J. Ghoshal, R. Laskar, and D. Pillone, "Further results on minimal rankings", *Ars Combin.* **52** (1999), 181–198. MR 2000f:05036 Zbl 0977.05048

[Hsieh 2002] S.-y. Hsieh, "On vertex ranking of a starlike graph", *Inform. Process. Lett.* **82**:3 (2002), 131–135. MR 2002k:05085 Zbl 1013.68141

[Isaak et al. 2009] G. Isaak, R. Jamison, and D. Narayan, "Greedy rankings and arank numbers", *Inform. Process. Lett.* **109**:15 (2009), 825–827. MR 2532182

[Jamison 2003] R. E. Jamison, "Coloring parameters associated with rankings of graphs", *Congr. Numer.* **164** (2003), 111–127. MR 2005d:05129 Zbl 1043.05049

[Kostyuk and Narayan $\geq$ 2010] V. Kostyuk and D. A. Narayan, "Minimal k-rankings for cycles", *Ars Combin.*. To appear.

[Kostyuk et al. 2006] V. Kostyuk, D. A. Narayan, and V. A. Williams, "Minimal rankings and the arank number of a path", *Discrete Math.* **306**:16 (2006), 1991–1996. MR 2007b:05094 Zbl 1101.05040

[Laskar and Pillone 2000] R. Laskar and D. Pillone, "Theoretical and complexity results for minimal rankings", *J. Combin. Inform. System Sci.* **25**:1-4 (2000), 17–33. MR 2001m:05244

[Laskar and Pillone 2001] R. Laskar and D. Pillone, "Extremal results in rankings", *Congr. Numer.* **149** (2001), 33–54. MR 2002m:05173 Zbl 0989.05058

[Leiserson 1980] C. E. Leiserson, "Area efficient graph layouts for VLSI", pp. 270–281 in *21st Ann. Symposium on Foundations of Computer Science* ($FOCS$), IEEE, 1980.

[Novotny et al. 2009a] S. Novotny, J. Ortiz, and D. Narayan, "Maximum minimal rankings of oriented trees", *Involve* **2**:3 (2009), 289–295. MR 2551126 Zbl 1177.05044

[Novotny et al. 2009b] S. Novotny, J. Ortiz, and D. A. Narayan, "Minimal k-rankings and the rank number of P_n^2", *Inform. Process. Lett.* **109**:3 (2009), 193–198. MR 2009m:05067 Zbl 05721970

[Sen et al. 1992] A. Sen, H. Deng, and S. Guha, "On a graph partition problem with application to VLSI layout", *Inform. Process. Lett.* **43**:2 (1992), 87–94. MR 1187395 Zbl 0764.68132

[de la Torre et al. 1992] P. de la Torre, R. Greenlaw, and T. Przytycka, "Optimal tree ranking is in NC", *Parallel Process. Lett.* **2**:1 (1992), 31–41.

Received: 2009-03-25 Revised: 2010-04-15 Accepted: 2010-04-15

jpo208@lehigh.edu
Department of Mathematics, Lehigh University,
Bethlehem, PA 18015, United States

anz1206@rit.edu
School of Mathematical Sciences,
Rochester Institute of Technology, 85 Lomb Memorial Drive,
Rochester, NY 14623, United States

hking@callutheran.edu
Mathematics Department, California Lutheran University,
Thousand Oaks, CA 91360, United States

dansma@rit.edu
School of Mathematical Sciences,
Rochester Institute of Technology, 85 Lomb Memorial Drive,
Rochester, NY 14623, United States
http://people.rit.edu/~dansma/

mirko.hornak@upjs.sk
Institute of Mathematics, P.J. Šafá rik University,
Jesenná 5, 040 01, Košice, Slovakia

INVOLVE 3:2(2010)

An unresolved analogue of the Littlewood Conjecture

Clarice Ferolito

(Communicated by Nigel Boston)

This article begins with an introduction to a conjecture made around 1930 in the area of Diophantine approximation: the Littlewood Conjecture. The conjecture asks whether any two real numbers can be simultaneously well approximated by rational numbers with the same denominator. The introduction also focuses briefly on an analogue of this conjecture, regarding power series and polynomials with coefficients in an infinite field. Harold Davenport and Donald Lewis disproved this analogue of the Littlewood Conjecture in 1963. Following the introduction we focus on a claim relating to another analogue of this conjecture. In 1970, John Armitage believed that he had disproved an analogue of the Littlewood Conjecture, regarding power series and polynomials with coefficients in a finite field. The remainder of this article shows that Armitage's claim was false.

1. Introduction

Through studying the results of John Littlewood and Godfrey Hardy on topics of Diophantine approximation, Littlewood's student Donald Spencer questioned whether any two real numbers can be approximated simultaneously by rational numbers with the same denominator. For some reason this conjecture was attributed to Littlewood and is known as *Littlewood's problem of Diophantine approximation* [Burkill 1979], or simply the *Littlewood Conjecture*.

To state it more formally, we fix some notation. As usual, $|n|$ denotes the absolute value of a number n. For x a real number, let $\|x\|$ denote the Euclidean distance of x to the nearest integer: $\|x\| = \inf_{a\in\mathbb{Z}} |x-a|$.

Conjecture 1.1 (Littlewood Conjecture). *For every θ, $\phi \in \mathbb{R}$ and for all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that*

$$n\,\|n\theta\|\,\|n\phi\| < \varepsilon.$$

MSC2000: 11K60.

Keywords: Littlewood Conjecture, John Vernon Armitage.

No one has been able to prove this, in part because the method of continued fractions commonly used with approximations cannot be used for simultaneous approximations. The following definitions will aid in describing the analogue of the Littlewood Conjecture for power series and polynomials, which is easier to study than the original conjecture.

For any field K, let $K[t]$ denote the set of polynomials with coefficients in K. Define the norm of $N \in K[t]$ as $|N|_K = e^h$, where h is the degree of N. Series with coefficients in K, possibly infinitely many negative exponents, and finitely many positive exponents form the field $K((t^{-1}))$. For every $\Psi \in K((t^{-1}))$, define the norm of Ψ to be $\|\Psi\|_K = e^l$, where l is the greatest negative exponent of t. For example, if $K = \mathbb{R}$ and $\Psi(t) = 12t^{50} + 3t^9 + 2 + 5t^{-11} + 20t^{-99} + \cdots \in K((t^{-1}))$, then $\|\Psi(t)\|_K = e^{-11}$.

Conjecture 1.2 (Polynomial analogue of the Littlewood Conjecture). *Let K be a field and consider Θ, $\Phi \in K((t^{-1}))$. For every $\varepsilon > 0$, there exists $N \in K[t]$ such that*

$$|N|_K \|N\Theta\|_K \|N\Phi\|_K < \varepsilon.$$

Davenport and Lewis [1963] proved that this analogue fails when K is an infinite field. Baker [1964] furthered this result by showing that $e^{1/t}$ and $e^{2/t} \in K((t^{-1}))$ serve as counterexamples to the analogue of the Littlewood Conjecture when K is the set of real numbers. With the analogue of the Littlewood Conjecture settled when K is an infinite field, the next problem to solve is the analogue with K a finite field.

2. Armitage's claim

Armitage [1970] published a corrigendum and addendum to his article from the previous year, entitled *An analogue of a problem of Littlewood* [Armitage 1969]. At first, it appeared that Armitage had disproved the analogue of the Littlewood Conjecture when K is a finite field of characteristic greater than or equal to 5. For many years, mathematicians accepted this claim. Armitage's proof appeared to imitate Baker's proof for his counterexample to the analogue of the Littlewood Conjecture with $K = \mathbb{R}$.

However, we found a parenthetical comment in [Adamczewski and Bugeaud 2007] that Armitage's counterexample does not hold, an observation these authors attribute to Bernard de Mathan. We also found a reference in [Larcher and Niederreiter 1993] that Yves Taussat, a student of Mathan, disproved Armitage's claim in his Ph.D. thesis [Taussat 1986]. However, in this paper, Taussat did not show why Armitage's counterexample fails. Below we provide a simplified wording of

Armitage's claim and in the next section we will show that his claim fails to disprove the analogue of the Littlewood Conjecture for K a finite field of characteristic $p \geq 5$.

Claim 2.1 [Armitage 1970]. *Let K be a field of characteristic $p > 3$. Define the norm of $N \in K[t]$ as $|N|_K = p^{\deg N}$, and the norm of $\Psi \in K((t^{-1}))$ as $\|\Psi\|_K = p^l$, where l is the greatest negative exponent of t in Ψ. Define $\Theta, \Phi \in K((t^{-1}))$ by*

$$\Theta(t) = (1+t^{-1})^{1/3}, \quad \Phi(t) = (1+t^{-1})^{2/3}.$$

Then for all nonzero $N \in K[t]$,

$$|N|_K \, \|N\Theta\|_K \, \|N\Phi\|_K \geq p^{-17}.$$

Note that Armitage uses p for the base of the norms rather than e, so his "lower bound", p^{-17}, for $|N|_K \, \|N\Theta\|_K \, \|N\Phi\|_K$ specifies the characteristic of K.

To show that Armitage's claim fails, we prove the following theorem.

Theorem 2.2. *Let K be a field of characteristic $p > 3$. Define $\Theta, \Phi \in K((t^{-1}))$ by*

$$\Theta = (1+t^{-1})^{1/3}, \quad \Phi = (1+t^{-1})^{2/3}.$$

Given any $\varepsilon > 0$, there exists a polynomial $N \in K[t]$ such that

$$|N|_K \, \|N\Theta\|_K \, \|N\Phi\|_K < \varepsilon.$$

The proof is divided into two cases, depending on the residue of p modulo 3.

3. Preliminary lemmas

Lemma 3.1. *For any prime p congruent to 2 modulo 3 and not equal to 2, the coefficient of t^{-n} in the expansion of $(1+t^{-1})^{1/3}$ is congruent to 0 modulo p if $(p^3+1)/3 < n < p^3$.*

Proof of Lemma 3.1. Consider the number of factors of p in the numerator of the coefficient of t^{-n} in $\Theta = (1+t^{-1})^{1/3}$. By the binomial theorem, this coefficient is

$$\binom{\frac{1}{3}}{n} = \frac{(-1)^{n-1}\big(1(3\cdot 1-1)(3\cdot 2-1)\cdots(3(n-1)-1)\big)}{3^n\, n!}.$$

Thus, the last term in the numerator of the coefficient of t^{-n} is $3n-4$. Since 3 always has a multiplicative inverse modulo p, we have

$$3l-1 \equiv 3m-1 \pmod{p} \iff l \equiv m \pmod{p}.$$

Moreover, the greatest common divisor of 3 and p^2 is 1 for any $p \neq 3$, so

$$3l-1 \equiv 3m-1 \pmod{p^2} \iff l \equiv m \pmod{p^2}.$$

Since $3l-1 \equiv 3m-1 \pmod{p} \iff l \equiv m \pmod{p}$, p divides at least $\lfloor n/p \rfloor$ terms in the numerator of the coefficient of t^{-n}. Similarly, because

$$3l-1 \equiv 3m-1 \pmod{p^2} \iff l \equiv m \pmod{p^2},$$

p^2 divides at least $\lfloor n/p^2 \rfloor$ terms.

Since $p \equiv 2 \pmod 3$, p^3 will divide a term in the numerator of a coefficient if and only if for some $r \in \mathbb{Z}^+$ $(0 \le r \le n-1)$ there exists $s \in \mathbb{Z}$ such that $3r-1 = sp^3$. In other words, $r = (sp^3+1)/3$ must be a positive integer. Thus, such an s must satisfy $s \equiv 1 \pmod 3$. We are only considering $n \le p^3-1$ and $r = (sp^3+1)/3 \le n-1$, so $(sp^3+1)/3 \le p^3-2$. Thus, sp^3 will divide a numerator in the coefficient of t^{-n} for $n \le p^3-1$ if and only if $s \le (3p^3-7)/p^3 < 3$ and $s \equiv 1 \pmod 3$. In other words, the only term that could be divisible by p^3 in the numerator of the coefficient of t^{-n} for $n \le p^3-1$ is p^3 itself. The final term in the numerator of the coefficient of $t^{-(p^3+4)/3}$ is $3((p^3+4)/3-1)-1 = p^3$. Thus, the first time that p^3 appears in the numerator of a coefficient of t^{-n} is actually when $n = (p^3+4)/3$. So, for each $(p^3+4)/3 \le n \le p^3-1$, the numerator of the coefficient of t^{-n} has exactly one term that is divisible by p^3. This and the preceding paragraph show that for each $(p^3+4)/3 \le n \le p^3-1$, the numerator of the coefficient of t^{-n} has at least $\lfloor n/p \rfloor + \lfloor n/p^2 \rfloor + 1$ factors of p.

Now looking at the denominator of the coefficient of t^{-n} for $(p^3+4)/3 \le n \le p^3-1$, the only powers of p that divide $n!$ are p and p^2. So there are only $\lfloor n/p \rfloor + \lfloor n/p^2 \rfloor$ factors of p in the denominator of t^{-n} for $(p^3+4)/3 \le n \le p^3-1$.

Therefore, for any $(p^3+4)/3 \le n \le p^3-1$, the numerator of the coefficient of t^{-n} will have at least one more factor of p than the denominator and the coefficient will be congruent to zero modulo p. □

The proof of the next lemma is similar to that of Lemma 3.1.

Lemma 3.2. *For any prime p congruent to* 1 *modulo* 3, *the coefficient of t^{-n} in the expansion of $(1+t^{-1})^{2/3}$ is congruent to zero modulo p if $(p^2+2)/3 < n < p^2$.*

Lemma 3.3. *For any prime $p > 3$ and any even positive integer b, there exists an integer $a < 0$ such that $\frac{1}{3} = a + p^b \cdot \frac{1}{3}$.*

Proof of Lemma 3.3. We have

$$\tfrac{1}{3} = a + p^b \cdot \tfrac{1}{3} \iff a = (1-p^b)/3 \in \mathbb{Z} \iff p^b \equiv 1 \pmod 3.$$

For any $p > 3$, $p \equiv 1 \pmod 3$ or $p \equiv 2 \pmod 3$. Obviously, $1^2 \equiv 1 \pmod 3$, but also $2^2 \equiv 1 \pmod 3$. Thus, for any prime $p > 3$, $p^2 \equiv 1 \pmod 3$. This further implies that $p^{2k} = (p^2)^k \equiv 1^k \equiv 1 \pmod 3$ for any $k \in \mathbb{N}$. Therefore, we have shown that for any even integer b, $p^b \equiv 1 \pmod 3$. □

The proof of the following lemma is analogous to that of Lemma 3.3.

Lemma 3.4. *For any prime $p > 3$ and any even positive integer b, there exists an integer $a < 0$ such that $\frac{2}{3} = a + p^b \cdot \frac{2}{3}$.*

4. Proof of Theorem 2.2

By assumption, the characteristic of K is $p > 3$.

First case: $p \equiv 2 \pmod 3$. By Lemma 3.3, for b an even positive integer, there exists a negative integer a such that $(1+t^{-1})^{1/3} = (1+t^{-1})^{a+p^b/3}$. Multiplying both sides by $(1+t^{-1})^{-a}$, yields $(1+t^{-1})^{-a}\,(1+t^{-1})^{1/3} = (1+t^{-1})^{p^b/3}$. Since we are working in a field with characteristic p, $(1+t^{-1})^{p^b/3} = (1+t^{-p^b})^{1/3}$, and therefore

$$(1+t^{-1})^{-a}\,(1+t^{-1})^{1/3} = (1+t^{-p^b})^{1/3}.$$

Multiplying both sides by t^{-a} results in

$$(1+t)^{-a}\,(1+t^{-1})^{1/3} = t^{-a}(1+t^{-p^b})^{1/3}.$$

Now applying Lemma 3.1, we know that the coefficient of t^{-i} in $t^{-a}(1+t^{-p^b})^{1/3}$ is congruent to zero modulo p for each i with $a+((p^3+1)/3)p^b < i < a+(p^3)p^b$. Multiplying

$$(1+t)^{-a}\,(1+t^{-1})^{1/3} = t^{-a}(1+t^{-p^b})^{1/3}$$

by $t^{a+((p^3+1)/3)p^b}$, we have

$$t^{a+((p^3+1)/3)p^b}(1+t)^{-a}\,(1+t^{-1})^{1/3} = q(t) + c\,t^{p^b((-2p^3+1)/3)} + \cdots,$$

where $q(t) \in K[t]$ and $c \not\equiv 0 \pmod p$.

Let N be the polynomial

$$N(t) = t^{a+((p^3+1)/3)p^b}(1+t)^{-a}.$$

Then by the definitions for the norm of a polynomial and the norm of a power series,

$$|N| = p^{((p^3+1)/3)p^b} \quad \text{and} \quad \|N\Theta\| = p^{\,p^b((-2p^3+1)/3)}.$$

Therefore

$$|N|\,\|N\Theta\| = p^{\,p^b((2-p^3)/3)}.$$

For any $\varepsilon > 0$, choosing an even positive integer b such that

$$b > \log_p\left|\frac{\log_p(\varepsilon)}{((2-p^3)/3)}\right|$$

implies that $|N|\,\|N\Theta\| < \varepsilon$.

Second case: $p \equiv 1 \pmod 3$. A similar method of proof works in this case, if we choose as the approximating polynomial

$$N = t^{a+((p^2+2)/3)p^b}(1+t)^{-a},$$

with b an even positive integer such that

$$b > \log_p \left| \frac{\log_p(\varepsilon)}{((4-p^2)/3)} \right|.$$

Since $\|N\Psi\| \le p^0$ for any $\Psi \in K((t^{-1}))$, together these cases show that Theorem 2.2 holds with the approximating polynomial N chosen as above depending on the characteristic of the finite field. □

Thus, Armitage's counterexample does not settle the analogue of the Littlewood Conjecture when K is a finite field of characteristic $p \ge 5$.

References

[Adamczewski and Bugeaud 2007] B. Adamczewski and Y. Bugeaud, "On the Littlewood conjecture in fields of power series", pp. 1–20 in *Probability and Number Theory* (Kanazawa, 2005), edited by S. Akiyama et al., Adv. Stud. Pure Math. **49**, Math. Soc. Japan, Tokyo, 2007. MR 2009e:11130 Zbl 05286791

[Armitage 1969] J. V. Armitage, "An analogue of a problem of Littlewood", *Mathematika* **16** (1969), 101–105. MR 42 #1768a Zbl 0188.35002

[Armitage 1970] J. V. Armitage, "Corrigendum and addendum: "An analogue of a problem of Littlewood".", *Mathematika* **17** (1970), 173–178. MR 42 #1768b

[Baker 1964] A. Baker, "On an analogue of Littlewood's Diophantine approximation problem", *Michigan Math. J.* **11** (1964), 247–250. MR 29 #2218 Zbl 0218.10052

[Burkill 1979] J. C. Burkill, "John Edensor Littlewood", *Bull. London Math. Soc.* **11**:1 (1979), 59–103. MR 80h:01027 Zbl 0409.01010

[Davenport and Lewis 1963] H. Davenport and D. J. Lewis, "An analogue of a problem of Littlewood", *Michigan Math. J.* **10** (1963), 157–160. MR 27 #4794 Zbl 0107.04202

[Larcher and Niederreiter 1993] G. Larcher and H. Niederreiter, "Kronecker-type sequences and non-Archimedean Diophantine approximations", *Acta Arith.* **63**:4 (1993), 379–396. MR 94c:11063

[Taussat 1986] Y. Taussat, *Approximation diophantienne dans un corps de séries formelles*, Ph.D. thesis, Université de Bordeaux, 1986.

Received: 2009-09-02 Revised: 2010-04-05 Accepted: 2010-06-02

cefero09@gmail.com

College of the Holy Cross, 355 California Street,
Newton, MA 02548, United States

Mapping the discrete logarithm

Daniel Cloutier and Joshua Holden

(Communicated by Carl Pomerance)

The discrete logarithm is a problem that surfaces frequently in the field of cryptography as a result of using the transformation $x \mapsto g^x \bmod n$. Analysis of the security of many cryptographic algorithms depends on the assumption that it is statistically impossible to distinguish the use of this map from the use of a randomly chosen map with similar characteristics. This paper focuses on a prime modulus, p, for which it is shown that the basic structure of the functional graph produced by this map is largely dependent on an interaction between g and $p-1$. We deal with two of the possible structures, permutations and binary functional graphs. Estimates exist for the shape of a random permutation, but similar estimates must be created for the binary functional graphs. Experimental data suggest that both the permutations and binary functional graphs correspond well to the theoretical predictions.

1. Introduction

Just a few decades ago, cryptography was considered a domain exclusive to national governments and militaries. However, the computer explosion has changed that. Every day, millions of people trust that their privacy will be protected as they make online purchases or communicate privately with a friend. Many of the cryptographic algorithms they will use are built upon a common transformation, namely

$$x \mapsto g^x \bmod n \tag{1}$$

where $\gcd(g,n) = 1$ and the transformation is considered as a function from $\{1,\dots,n-1\}$ to itself. (We will call functions of this form discrete exponentiation maps.) For instance, Diffie–Hellman key exchange [Diffie and Hellman 1976], RSA [Rivest et al. 1978], and the Blum–Micali pseudorandom bit generator [Blum and Micali 1984] all use discrete exponentiation maps. In particular, if n is a prime and g is a primitive root modulo that prime, then a discrete exponentiation map has

MSC2000: primary 11Y16; secondary 11-04, 94A60, 05A15.

Keywords: discrete logarithm problem, random map, functional graph.

Holden was supported in part by a Rose–Hulman Summer Professional Development Grant during the summer of 2009.

an inverse which is known as the *discrete logarithm*. The security of the Diffie–Hellman protocol and the Blum–Micali generator both rely on the idea that the discrete logarithm is difficult to calculate.

Furthermore, the analyses of the security of many algorithms rely on the idea that not only is calculating the inverse of a discrete exponentiation map difficult, but in fact that certain properties of discrete exponentiation maps and/or discrete logarithms cannot be predicted better than a random guess. (It is not known to the authors who first suggested this general idea; it may be folklore.) For example, in [Blum and Micali 1984] the cryptographic security of a particular pseudorandom bit generator relies on the hypothesis that a certain property of discrete exponentiation cannot be predicted better than a random guess. Similarly, [Boneh 1998] shows that if certain statistical properties of the Diffie–Hellman problem cannot be guessed better than randomly then the Diffie–Hellman protocol can be made much more efficient than otherwise. This paper will consider some statistics of maps on $\{1, \dots, n-1\}$ such that the expected values of these statistics for a randomly chosen map in a class containing the discrete exponentiation maps can be calculated theoretically. We conjecture that the particular values of these statistics for discrete exponentiation maps will resemble the expected values for the random maps. Furthermore, we will collect experimental data on discrete exponentiation maps for various values of g and n and compare them to our expected values to give evidence for this conjecture.

Some readers might be familiar with other papers that look at the discrete exponentiation map from a statistical point of view, such as [Canetti et al. 2000]. In both cases n is fixed and "measurements" are taken from a (nonrandom) sample which is derived from discrete exponentiation maps. The distribution of the measurements on the sample is then compared with the distribution of measurements taken from random samples of a certain population. In [Canetti et al. 2000], g is fixed, the measurements are a specified set of bits from triples of numbers, the sample is triples of the form (g^x, g^y, g^{xy}) (with varying x and y), and the population is all strings of bits. In this paper, the measurements are various graph-theoretic properties of functional graphs (as defined below). The sample is maps of the form (1) (with varying g) which have a certain property on their in-degrees and the population is all functional graphs with that same property.

2. Terminology and background

Throughout this paper, ϕ denotes the Euler phi function. The letter n will stand for an odd prime. We will examine mappings

$$f : S = \{1, 2, \dots, p-1\} \to S$$

of the form $x \mapsto g^x \bmod p$, where $p \geq 3$ is a prime modulus and $\gcd(g, p) = 1$. In some instances, it will prove to be useful to interpret the mappings as functional graphs. A functional graph is a directed graph such that each vertex must have exactly one edge directed out from it. The relationship between the mappings which interest us and functional graphs is straightforward. Each element in S can be interpreted as a vertex. The edges are defined such that an edge $\langle a, b\rangle$ is in the graph if and only if $f(a) = b$.

There are a number of statistics of interest derived from functional graphs; in particular, Flajolet and Odlyzko — henceforth abbreviated FO — have treated random mappings in detail. Following the conventions in [FO 1990b], let $f : S \to S$ be the transition function so that the edges in the functional graph can be expressed as the ordered pair $\langle x, f(x)\rangle$ for $x, f(x) \in S$. By applying the pigeonhole principle and noting that the cardinality of S is $p - 1$ we can say that by starting at any random point u_0 and following the sequence $u_1 = f(u_0)$, $u_2 = f(u_1), \ldots,$ there must be a $u_i = u_j$ after at most p iterations. Suppose u_i occurs before u_j in the sequence of nodes. In this case, the tail length is the number of iterations of the function from u_0 to u_i. The cycle length is the number of iterations from u_i to u_j. In more natural graphical terms, the tail length is the number of edges involved in the directed path from u_0 to u_i, and the cycle length is the number of edges (or equivalently nodes) involved in the directed path from u_i to itself. Additionally, a terminal node is one with no preimage, or more formally, x is a terminal node if $f^{-1}(x) = \varnothing$. A node is an image node if it is not a terminal node. Since each node has an out-degree of exactly one, each cycle with the trees grafted onto its nodes will form a connected component.

When a functional graph is produced from a discrete exponentiation function, we will call it a discrete exponentiation functional graph. The value of g plays a major role in determining the basic structure of discrete exponentiation functional graphs. In fact, as Theorem 1 formalizes, the interaction between g and $p - 1$ will effectively fix the in-degrees of the nodes in the graph. First, though, define an m-ary functional graph to be a graph where each node has in-degree of exactly zero or m. The proof of the following theorem is then straightforward.

Theorem 1. *Let p be fixed and let m be any positive integer that divides $p - 1$. Then as g ranges from 1 to $p-1$, there are $\phi((p-1)/m)$ different functional graphs which are m-ary produced by maps of the form $f : x \mapsto g^x \bmod p$. Furthermore, if r is any primitive root modulo p, and $g \equiv r^a \mod p$, then the values of g that produce an m-ary graph are precisely those for which $\gcd(a, p - 1) = m$.*

Theorem 1 gives a strong indication that the graphs generated by (1) have to be considered separately for different values of m. It should be noted, though, that there are some values of m which lead to completely predictable graphs. For instance, there is one $(p-1)$-ary graph that corresponds to $g \equiv 1 \mod p$. There is also one $((p-1)/2)$-ary graph that corresponds to $g \equiv -1 \mod p$. In general, however, an m-ary discrete exponentiation functional graph is not trivially predictable. This paper will restrict its focus to unary functional graphs (which will be referred to as permutations since they simply permute the numbers $1, \dots, p-1$) and binary functional graphs. As a consequence of Theorem 1, we can observe that the values of g which produce a permutation are precisely those which are primitive roots modulo p, and the values of g which produce a binary functional graph are precisely those which are the squares of primitive roots modulo p.

In cryptography, it is common to look for primes where $p-1$ has at least one large prime factor. For instance, the pseudorandom bit generator described in [Gennaro 2005], which is a modification of the Blum–Micali generator mentioned in Section 1, specifically requires the modulus to be of the form $p = 2q+1$ where q is also prime. A prime of this form is known as a *safe prime* (q is also known as a *Sophie Germain prime*). These primes are of interest here not only because of their extensive use in cryptography, but also because $p-1$ has only four divisors, namely 1, 2, $q = (p-1)/2$ and $2q = p-1$. In addition to the one $(p-1)$-ary and one $((p-1)/2)$-ary graph mentioned above, there are $\phi(q)$ permutations and $\phi(q)$ binary functional graphs which represent the remaining values of g (since $\phi(q)$ is $q-1$). Thus, not only do safe primes provide large numbers of permutations and binary functional graphs, but every graph generated by a safe prime is either trivial (the graphs where g is either 1 or -1) or fits into the theoretical framework presented in Section 3.

We can now present the central conjecture of this paper, which as far as we know has not been previously considered in this form:[1]

Conjecture 2. *The average values of the following statistics are asymptotically the same for m-ary discrete exponential functional graphs on $n = p-1$ nodes and for random m-ary functional graphs on n nodes as n goes to infinity*:

Number of components
Number of tail nodes
Number of image nodes
Average cycle length
Maximum tail length
(as seen from a random node)

Number of cyclic nodes
Number of terminal nodes
Maximum cycle length
Average tail length
(as seen from a random node)

We are a long way from proving this conjecture but we will give some supporting evidence for it in the cases of $m = 1$ and $m = 2$.

[1] Pollard [1978] considers functional graphs corresponding to a similar map when analyzing his kangaroo method. That map takes $x \mapsto xg^{f(x)}$ for some pseudorandom function f, however.

3. Theoretical results

In Theorem 1, it is shown that the in-degree of each node is dependent on the value of both g and p. This is clearly imposing a structure on any functional graphs generated using (1). While most of the parameters that are of interest depend on the exact graph generated, the number of image nodes can be computed directly from the values of g and p. The proof is again straightforward.

Theorem 3. *The number of image nodes in any m-ary graph is $(p-1)/m$.*

This fact helps quantify the repercussions of Theorem 1 and the restrictions on in-degree in m-ary graphs. The number of image nodes is a direct function of m which can greatly limit the shapes each graph can take on. None of the other parameters appear to be strictly controlled by m in this fashion.

3.1. ***Permutations.*** Predicting the behavior of the permutations is, in many ways, much easier than other m-ary graphs. The most important reason for this is that there are no terminal nodes or tail nodes. This follows quickly from the definition of a permutation as a unary functional graph and the fact that the sum of the in-degrees must be the same as the sum of the out-degrees. Each node has an out-degree of exactly one, and if any node were to have an in-degree of zero, then, by the pigeon-hole principle, at least one node must have an in-degree of more than one. This is not allowed so each node must have in-degree of exactly one. Furthermore, since every tail must contain at least one terminal node, this also implies that every node is cyclic. The parameters that can then be determined from the definition of a permutation are:

Number of cyclic nodes	n	Number of tail nodes	0
Number of terminal nodes	0	Number of image nodes	n
Average tail length	0	Maximum tail length	0

Theorem 4. *The expected values for the number of components, the average cycle length as seen from a random node and the maximum cycle length in a random permutation of size n have the following asymptotic forms*:

$$\textit{Number of components} = \sum_{i=1}^{n} \frac{1}{i} + o(\log n), \tag{i}$$

$$\textit{Average cycle length} = \frac{n+1}{2} + o(1), \tag{ii}$$

$$\begin{aligned}\textit{Maximum cycle length} &= n \int_0^\infty \left[1 - \exp\left(-\int_v^\infty e^{-u} \frac{du}{u}\right)\right] dv + o(n) \\ &\approx 0.62432965n + o(n).\end{aligned} \tag{iii}$$

Parts (i) and part (ii) are fairly well known. Part (iii) seems to have first been solved in [Shepp and Lloyd 1966]. An alternative solution and proof more similar to the methods used here is offered in [FO 1990a].

3.2. ***Binary functional graphs.*** While estimates for the parameters investigated here exist in the literature for the random functional graphs and permutations, it does not appear that estimates for binary functional graphs have ever been all collected in one place. However, the methods in [FO 1990b] can be extended to develop these estimates, and some of the following results have appeared in various places already. Imitating those methods, we first need to convert our ideas of a binary functional graph into corresponding generating functions. We first note that a binary functional graph is a set of components. Each component is a cycle of nodes with each node having an attached binary tree to bring its in-degree to two. A binary tree is either a node (terminal node) or a node with two binary trees attached. Finally, a node is simply an atomic unit. A moment's reflection should indicate that this natural specification does, in fact, specify a binary functional graph.[2]

Imitating the transformations in [FO 1990b, Section 2.1], the generating functions of interest are

$$f(z) = e^{c(z)} = \frac{1}{1 - zb(z)}, \tag{2}$$

$$c(z) = \ln \frac{1}{1 - zb(z)}, \tag{3}$$

$$b(z) = z + \frac{1}{2} z b^2(z). \tag{4}$$

Here f generates the number of binary functional graphs, c generates the number of components, and b generates the number of binary trees of a given size. Solving the quadratic formula for (4), we can produce the following formulas for f and c which simplify some of the cases:

$$f(z) = \frac{1}{\sqrt{1-2z^2}}, \qquad c(z) = \ln \frac{1}{\sqrt{1-2z^2}} \tag{5}$$

See also [FO 1990b, (70)] and [Flajolet et al. 1991, Theorem 11].

To compute asymptotic forms of any of the statistics of interest, we must first compute an asymptotic form for f to normalize results. The following derivations

[2] In the notation of [FO 1990b]:

BinFunGraph	= set(Components),
Component	= cycle(Node*BinaryTree),
BinaryTree	= Node + Node*set(BinaryTree, cardinality = 2),
Node	= Atomic Unit.

give only a highlight of the methods used by Flajolet and Odlyzko. The interested reader is encouraged to see [FO 1990a; 1990b] for detailed proofs.

From the formula for $f(z)$ in (5) it is clear that there is a singularity at $z = 1/\sqrt{2}$. Performing a singularity analysis[3] as in [FO 1990b, Section 2], the asymptotic form for f falls out quickly as

$$f(z) \sim \frac{2^{n/2}}{\sqrt{\pi n/2}}. \tag{6}$$

In at least one case, there are some important second-order interactions between the error terms of the number of graphs and the appropriate statistic. In these cases, a more exact form of (6) must be used. Expanding one more term in the expansion of f gives

$$f(z) \sim \frac{2^{n/2}}{\sqrt{\pi n/2}} - \frac{2^{n/2}}{4n\sqrt{\pi n/2}} = \frac{2^{n/2}(4n-1)}{4n\sqrt{\pi n/2}}. \tag{7}$$

In most cases, using this more precise expansion of f is not necessary and does not change the results. Therefore, in all but the necessary cases, (6) will be used.

We begin by deriving the results for the simplest parameters.

Theorem 5. *The expected values for the number of components, number of cyclic nodes, number of tail nodes, number of terminal nodes and number of image nodes in a random binary functional graph of size n, as $n \to \infty$ have the following asymptotic forms*:

$$\textit{Number of components} = \frac{\ln(2n) + \gamma}{2} + o(1), \tag{i}$$

$$\textit{Number of cyclic nodes} = \sqrt{\pi n/2} - 1 + o(1), \tag{ii}$$

$$\textit{Number of tail nodes} = n - \sqrt{\pi n/2} + 1 + o(1), \tag{iii}$$

$$\textit{Number of terminal nodes} = n/2, \tag{iv}$$

$$\textit{Number of image nodes} = n/2. \tag{v}$$

In part (i), γ represents the Euler constant which is approximately 0.57721566. The results for parts (iv) and (v) can in fact be shown to be exact and not merely asymptotic. The highlights of the proofs as they differ from those in [FO 1990b] follow.

Proof. As in [FO 1990b], the following bivariate generating functions need to be defined with parameter u marking the elements of interest. The generating

[3] The analyses in this paper have been performed using the computer algebra program Maple and the packages created as part of the Algorithms Project at INRIA, Rocquencourt, France. The packages can be found online at http://algo.inria.fr/libraries/#down.

functions for the number of components, number of cyclic nodes and number of terminal nodes are respectively:

$$\xi_1(u,z) = \exp\left(u \ln \frac{1}{1-zb(z)}\right), \tag{8}$$

$$\xi_2(u,z) = \frac{1}{1-uzb(z)}, \tag{9}$$

$$\xi_3(u,z) = \frac{1}{\sqrt{1-2uz^2}}. \tag{10}$$

(Equation (9) may also be found in [Flajolet et al. 1991, Theorem 11].) Imitating the methods in [FO 1990b], the mean value generating function, $\Xi(z)$, is found by taking the partial derivative of $\xi(u,z)$ with respect to u and evaluating at $u = 1$. This yields

$$\Xi_1(z) = \frac{1}{1-zb(z)} \ln \frac{1}{1-zb(z)}, \tag{11}$$

$$\Xi_2(z) = \frac{zb(z)}{(1-zb(z))^2}, \tag{12}$$

$$\Xi_3(z) = \frac{z^2}{(1-2z^2)^{3/2}}. \tag{13}$$

The forms in the statement of the theorem follow by expanding around the singularity $z = 1/\sqrt{2}$, applying singularity analysis as in [FO 1990b], and normalizing parts (i) and (ii) by (6) and (iv) by (7). Parts (iii) and (v) follow from parts (ii) and (iv) respectively since the respective pairs must sum to n. Also note that part (iv) can also be derived in an elementary fashion from the definition of the binary functional graph. □

The asymptotic values for the average length of cycles and tails as seen from a random point in the graph are also interesting. The asymptotic forms of these values are given in Theorem 6.

Theorem 6. *The expected values for the cycle size and tail length as seen from a random node in a random binary functional graph of size n have the following asymptotic forms as $n \to \infty$:*

$$\textit{Average cycle length } = \sqrt{\pi n/8} + o(\sqrt{n}), \tag{i}$$

$$\textit{Average tail length } = \sqrt{\pi n/8} + o(\sqrt{n}). \tag{ii}$$

Proof. In order to calculate the average cycle length and average tail length, the generating functions must be manipulated to account for each node in the cycle or tail. This can be done by using the same methods as in the previous proof, but marking only one component or tail at a time. This is essentially the same as the strategy which is used to prove the result for average cycle size in [FO 1990b]. More background on the method can be found there.

Let $\xi_1(z)$ be the exponential generating function for the total cycle length over all binary functional graphs and $\xi_2(z)$ be the exponential generating function for the total tail length. Then, $\xi_1(z)$ can be defined as

$$\xi_1(z) = \frac{\partial^2}{\partial w \partial u}\left[\frac{1}{1-\sqrt{1-2z^2}} \ln\left(\frac{1}{1-u\left(1-\sqrt{1-2(zw)^2}\right)}\right)\right]_{u=1,w=1}. \tag{14}$$

In (14), u marks the cyclic nodes in the component we are considering and w marks all nodes in that component, so that each node in the component is weighted with the number of nodes in the cycle. (In [Salvy 1997], the method of "decorated" graphs is used to develop a generating function for a variation of this problem.)

In order to compute total tail length we need a version of the generating function for binary trees which marks the edges along one tail. We can write that as

$$\beta(z,u) = z + \tfrac{1}{2}zb^2(z) + uzb(z)\beta(z,u). \tag{15}$$

Then solving (15) and plugging it in appropriately gives us

$$\xi_2(z) = \frac{\partial}{\partial u}\left[\frac{1}{\sqrt{1-2z^2}} \frac{1}{\sqrt{1-2z^2}} \frac{u(1-\sqrt{1-2z^2})}{\left(1-u(1-\sqrt{1-2z^2})\right)}\right]_{u=1}. \tag{16}$$

Note that the first factor in (16) is for the unmarked components and the second is for the unmarked trees in the marked components. (In [Salvy 1997] and [Flajolet et al. 1989; Mishna 2004],[4] the methods of "decorated" graphs and attribute grammars, respectively, are used to develop the same generating function.)

Performing a singularity analysis of the two generating functions and normalizing by $2^{n/2}/(n\sqrt{\pi n/2})$, as done in the previous theorems, leads to the statement of the theorem. The additional factor of n in the denominator is needed to compensate for the fact that the parameters were estimated across all nodes in the graph and the goal is to determine them from any single random node in the graph. □

The final parameters that needs to be calculated are the average maximum cycle length and the average maximum tail length.

Theorem 7. *The expected sizes of the largest cycle and the largest tail in a random binary functional graph of size n have the following asymptotic forms as $n \to \infty$:*

[4] According to [Flajolet et al. 1989], results from this analysis were first obtained by hand in [Flajolet 1979].

$$\begin{aligned} \textit{Largest cycle} &= \sqrt{\frac{\pi n}{2}} \int_0^\infty \left[1 - \exp\left(-\int_v^\infty e^{-u} \frac{du}{u}\right)\right] dv + o(\sqrt{n}) && \text{(i)} \\ &\approx 0.78248\sqrt{n} + o(\sqrt{n}); \\ \textit{Largest tail} &= \sqrt{2\pi n}\ln 2 - 3 + 2\ln 2 + o(1) && \text{(ii)} \\ &\approx 1.73746\sqrt{n} - 1.61371 + o(1). \end{aligned}$$

Proof. The proof for part (i) result follows precisely the methods of [FO 1990b] with substitution of the proper generating function f, and is therefore omitted.

The proof for part (ii) follows a combination of [FO 1990b, Theorem 6] and [FO 1982, Sections 3–5]. Let $b^{[h]}(z)$ be the exponential generating function for the number of binary trees with height at most h and $f^{[h]}(z)$ be the exponential generating function for the number of binary functional graphs with maximum tail length less than or equal to h, so that (as in [FO 1990b, Equations 41 and 42])

$$f^{[h]}(z) = \frac{1}{1 - zb^{[h]}(z)}$$

and

$$b^{[h+1]}(z) = z + \tfrac{1}{2}z(b^{[h]}(z))^2, \quad b^{[0]}(z) = z.$$

Now, as in [FO 1982, Proposition 2], note that

$$b(z) - b^{[h+1]}(z) = \tfrac{1}{2}z(b(z) - b^{[h]}(z))(b(z) + b^{[h]}(z)),$$

so if we let

$$e_h(z) = \frac{b(z) - b^{[h]}(z)}{2b(z)},$$

then

$$e_{h+1}(z) = (1 - \sqrt{1 - 2z^2})e_h(z)(1 - e_h(z)).$$

Now we want to approximate $e_h(z)$ with a function of h and some $\epsilon(z)$. If we let $\epsilon = \sqrt{1 - 2z^2}$ then we have

$$e_{j+1} = (1 - \epsilon)e_j(1 - e_j); \quad e_{-1} = 2.$$

This is essentially the same recursion as in [FO 1982]. As in Lemma 5 there, we can then “normalize” and “take inverses” to get the approximation

$$e_h \approx \frac{(1-\epsilon)^{h+1}\epsilon}{1 - (1-\epsilon)^{h+1}}. \tag{17}$$

The details of the error bounds proceed as in [FO 1982]; we omit them here.

The generating function associated to the average maximum tail length is, as in [FO 1990b, Equation 43],

$$\Xi(z) = \sum_{h\geq 0} \Big[\frac{1}{1-zb(z)} - \frac{1}{1-zb^{[h]}(z)} \Big],$$

and we proceed as in [FO 1990b, Equation 51] to write

$$\Xi(z) = \frac{2zb(z)}{1-zb(z)} \sum_{h\geq 0} \frac{e_h(z)}{1-zb(z)+2e_h(z)zb(z)}.$$

Putting this entirely in terms of ϵ and h, and shifting the index of summation for convenience, we can write

$$\Xi(z) \approx \frac{2(1-\epsilon)}{\epsilon} \sum_{h\geq 1} \frac{(1-\epsilon)^h}{1+(1-2\epsilon)(1-\epsilon)^h}. \tag{18}$$

We approximate the sum with an integral, using Euler–Maclaurin summation. Taking the integral and noting that $\ln(1-\epsilon) \sim -\epsilon$ as $\epsilon \to 0$, we finally get

$$\Xi(z) \approx \frac{2(1-\epsilon)}{\epsilon^2(1-2\epsilon)} \ln(2-3\epsilon+2\epsilon^2). \tag{19}$$

The next step is to substitute $\epsilon = \sqrt{1-2z^2}$ into (19) and do the singularity analysis, which gives us the statement of the theorem. □

4. Observed results

In [Holden 2002; Holden and Moree 2004; 2006], heuristics and observed values for the number of small cycles (fixed points and two-cycles) in discrete exponentiation graphs are given. Our methods build on this to generate experimental data for the parameters described by the theoretical predictions in Section 3. The method of data collection was straightforward. A prime was chosen as the modulus and then for each $g \in \{1, 2, 3, \dots, p-1\}$, the corresponding discrete exponentiation binary functional graph or permutation was generated. The results were then computed as average statistics over all $p-1$ graphs observed. The permutations and binary functional graphs were noted and their results were also tabulated separately. In this manner, the data can be examined in its complete form over all graphs and individually over the permutations and binary functional graphs. The generation and analysis of each of the graphs was handled by C++ code written by the first author.

	100043	100057	106261
Permutations	50020	30240	21120
Binary functional graphs	50020	15120	10560
Total functional graphs	100042	100056	106260

Table 1. The number of permutations, binary functional graphs and total discrete exponentiation functional graphs associated with $p = 100043$, $p = 100057$, and $p = 106260$.

The primes chosen for these calculations were

$$\begin{aligned} 100043 &= 2 \cdot 50021 + 1, \\ 100057 &= 2^3 \cdot 3 \cdot 11 \cdot 379 + 1, \\ 106261 &= 2^2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 23 + 1. \end{aligned}$$

The total number of graphs, permutations and binary functional graphs can be computed using Theorem 1 and are shown in Table 1.

In Section 4.1, the observed results for the discrete exponentiation permutations will be compared to the theoretical results for random permutations given in Theorem 4. Finally, the observed results for the discrete exponentiation binary functional graphs will be examined in Section 4.2. Theorems 5 through 7 will provide the theoretical predictions for these values on random binary functional graphs. Since the terminal nodes and tail nodes can be directly computed from the image nodes and cyclic nodes, including them in the collected data does not add any insight. For this reason, they have both been excluded from the analysis conducted in the following sections. The Appendix gives some of the interesting extremal data such as the longest cycle observed for each prime. More information on the data and how they were computed may be found in [Cloutier 2005].

4.1. ***Permutation results.*** The results of looking at only the values of g that were a primitive root modulo p (and thus produced permutation discrete exponentiation graphs) can be found in Table 2.

The percent error here is nearly zero in every instance. This seems to indicate that there are no obvious structural differences between a random permutation and a permutation generated by the process used here.

4.2. ***Binary functional graph results.*** The binary functional graphs should prove more interesting than the permutations examined in the previous section. Unlike permutations, binary functional graphs do not appear to have been previously studied in detail. The statistics derived from the binary discrete exponentiation functional graphs and the error when compared to the results derived for random binary functional graphs in Section 3.2 can be found in Table 3.

The number of image nodes came out exactly as expected and predicted by Theorem 3. However, in many other cases the results were nearly as good. The relative size of the error decreases as the number of binary discrete exponentiation functional graphs increases over the different primes. This is especially worth noting for $p = 100043$ which has over fifty thousand binary functional graphs while 100057 and 106261 have approximately fifteen thousand and ten thousand respectively. Since having more graphs appears to push the results closer to those derived in Section 3.2, this seems to further support the claim that any binary functional graph produced by our mapping does in fact resemble a randomly chosen binary functional graph.

5. Conclusions and future work

The transformation used here to generate functional graphs is an exceedingly important transformation in cryptography. If the output of the function were to fall into a predictable pattern, it could be an exploitable flaw in many algorithms considered secure today. For instance, the average cycle length seems particularly important for pseudorandom bit generators since, in many cases, it relates directly to the predictability of the pseudorandom bit generator. As Theorem 1 demonstrates, the use of (1) repeatedly forces a nontrivial structure onto the graphs generated. This is certainly worth investigating as any imposed structure may be open to an exploit.

The advantage of using a safe prime is that every nontrivial graph can be analyzed by the theoretical framework laid out in this paper. Their use is also very prevalent in cryptographic applications. As mentioned above, the pseudorandom bit generator specified in [Gennaro 2005] requires the use of a safe prime to defend against other attacks. However, the methods used for binary functional graphs in Section 3.2 can and should be extended to larger values of m. (This is currently underway in the case $m = 3$ and some results may be found in [Brugger and Frederick

	100043		100057		106261	
	Observed	Error	Observed	Error	Observed	Error
Components	12.081	0.083%	12.054	0.306%	12.126	0.205%
Avg cycle	49980.551	0.082%	50191.352	0.326%	53105.104	0.048%
Max cycle	62395.488	0.102%	62627.745	0.256%	66245.807	0.144%

Table 2. Observed results for the three primes over the permutation discrete exponentiation graphs, with corresponding errors.

	100043		100057		106261	
	Observed	Error %	Observed	Error %	Observed	Error %
Components	6.389	0.047	6.364	0.437	6.370	0.810
Cyclic nodes	395.303	0.029	395.858	0.105	408.433	0.217
Image nodes	50021	0	50028	0	53130	0
Avg cycle	198.319	0.056	197.766	0.230	202.651	0.795
Avg tail	197.961	0.125	197.550	0.339	202.422	0.907
Max cycle	247.261	0.094	247.302	0.082	256.986	0.754
Max tail	541.827	1.115	549.588	1.145	566.370	1.744

Table 3. The observed results for the three primes over all binary discrete exponentiation functional graphs generated and the corresponding percent errors.

2007; Brugger 2008].) In an ideal case, they should be extended to the general case of an m-ary graph, which can be specified as a set of components, each of which is a cycle of nodes with each node having an attached m-ary tree.[5] The associated generating functions for these functional graphs would be

$$f(z) = e^{c(z)}, \quad c(z) = \ln\left(1 - \frac{z}{(m-1)!}t^{m-1}(z)\right)^{-1}, \quad t(z) = z + \frac{z}{m!}t^{m}(z),$$

where $f(z)$ is the exponential generating function associated to the functional graphs, $c(z)$ is the exponential generating function associated to the connected components and $t(z)$ is associated to the trees. The methods in Section 3.2 could also be extended to obtain values for additional parameters such as the average and maximum tree size.

This paper has focused on the graphs generated when the modulus is prime. In practice, though, this is not always the case. For this reason, it could be worthwhile to attempt to extend the type of analysis done here to a composite modulus. Some work in this direction may be found in [Mace 2009].

While the data generated for this project appears to confirm that the graphs do tend toward the shape and structure of a random graph of the appropriate type, no data were collected on the distribution of the different parameters. This data could help to give a clearer picture of how closely individual graphs may be expected to exhibit the characteristics of a random graph, especially given the observation that primes with a larger number of binary functional graphs seem to conform better to

[5] In the notation of [FO 1990b]:

FunctionalGraph = set(Components),
Component = cycle(Node*Set(Tree, cardinality = $m-1$)),
Tree = Node + Node*set(Tree, cardinality = m),
Node = Atomic Unit.

prediction on the average. The methods used in [Flajolet et al. 1993] would seem to be potentially helpful here. In addition, finding the standard deviation for the parameters of interest across all graphs of the appropriate type would allow us to do a more sophisticated analysis of the observed errors. Initial work along these lines has been done for permutations in [Hoffman 2009] and for binary functional graphs in [Lindle 2008].

Appendix: Extremal data

For $p = 100043$, the longest cycle observed was 100042 which occurred for two different values of g. They were $g = 20812$ and $g = 94034$. The longest tail had a length of 1448 and was observed when $g = 89339$. There were five instances where the graphs contained no cycles longer than one which occurred for $g = 1$, 72116, 91980, 95997, and 100042.

The graphs generated by $p = 100057$ had an overall longest cycle of 100052 when $g = 58303$. The longest tail observed was 1589 when $g = 18115$. There were also 26 different values of g that produced a graph that did not have a cycle longer than one.

The largest cycle observed in graphs generated using $p = 106261$ was 106257 when $g = 102141$. The longest tail was 35822 when $g = 1480$. There were 92 different values of g that produced graphs with no cycles longer than a fixed point.

Acknowledgments

The authors thank the anonymous referee for advice on how to improve the clarity of the paper.

References

[Blum and Micali 1984] M. Blum and S. Micali, "How to generate cryptographically strong sequences of pseudorandom bits", *SIAM J. Comput.* **13**:4 (1984), 850–864. MR 86a:68021

[Boneh 1998] D. Boneh, "The decision Diffie–Hellman problem", pp. 48–63 in *Algorithmic number theory* (Portland, OR, 1998), edited by J. P. Buhler, Lecture Notes in Comput. Sci. **1423**, Springer, Berlin, 1998. MR 2000k:94024 Zbl 1067.94523

[Brugger 2008] M. F. Brugger, *Exploring the discrete logarithm with random ternary graphs*, senior thesis, Oregon State University, 2008, available at http://hdl.handle.net/1957/8777.

[Brugger and Frederick 2007] M. Brugger and C. Frederick, "The discrete logarithm problem and ternary functional graphs", *Rose-Hulman Undergraduate Mathematics Journal* **8**:2 (2007).

[Canetti et al. 2000] R. Canetti, J. Friedlander, S. Konyagin, M. Larsen, D. Lieman, and I. Shparlinski, "On the statistical properties of Diffie–Hellman distributions", *Israel J. Math.* **120**:part A (2000), 23–46. MR 2001k:11258 Zbl 0997.11066

[Cloutier 2005] D. R. Cloutier, *Mapping the discrete logarithm*, senior thesis, Rose-Hulman Institute of Technology, 2005, available at http://www.csse.rose-hulman.edu/images/docs/theses/DanielCloutier2005.pdf.

[Diffie and Hellman 1976] W. Diffie and M. E. Hellman, "New directions in cryptography", *IEEE Trans. Information Theory* **IT-22**:6 (1976), 644–654. MR 55 #10141

[Flajolet 1979] P. Flajolet, *Analyse d'algorithmes de manipulation d'arbres et de fichiers*, Ph.D. thesis, Université de Paris-Sud, Orsay, 1979.

[Flajolet et al. 1989] P. Flajolet, B. Salvy, and P. Zimmermann, "Lambda-upsilon-omega: The 1989 cookbook", Technical Report RR1073, Institut National de Recherche en Informatique et en Automatique, 1989, available at http://www.inria.fr/rrrt/rr-1073.html.

[Flajolet et al. 1991] P. Flajolet, B. Salvy, and P. Zimmermann, "Automatic average-case analysis of algorithms", *Theoret. Comput. Sci.* **79**:1, (Part A) (1991), 37–109. MR 92k:68049 Zbl 0768.68041

[Flajolet et al. 1993] P. Flajolet, Z. Gao, A. Odlyzko, and B. Richmond, "The distribution of heights of binary trees and other simple trees", *Combin. Probab. Comput.* **2**:2 (1993), 145–156. MR 94k:05061 Zbl 0795.05042

[FO 1982] P. Flajolet and A. Odlyzko, "The average height of binary trees and other simple trees", *J. Comput. System Sci.* **25**:2 (1982), 171–213. MR 84a:68056 Zbl 0499.68027

[FO 1990a] P. Flajolet and A. Odlyzko, "Singularity analysis of generating functions", *SIAM J. Discrete Math.* **3**:2 (1990), 216–240. MR 90m:05012 Zbl 0712.05004

[FO 1990b] P. Flajolet and A. M. Odlyzko, "Random mapping statistics", pp. 329–354 in *Advances in cryptology* (Houthalen, Belgium, 1989), edited by A. J. Menezes and S. A. Vanstone, Lecture Notes in Comput. Sci. **434**, Springer, Berlin, 1990. MR 1083961 Zbl 0747.05006

[Gennaro 2005] R. Gennaro, "An improved pseudo-random generator based on the discrete logarithm problem", *J. Cryptology* **18**:2 (2005), 91–110. MR 2007c:94124 Zbl 1084.68046

[Hoffman 2009] A. Hoffman, "Statistical investigation of structure in the discrete logarithm", *Rose-Hulman Undergrad. Math. J.* **10**:2 (2009).

[Holden 2002] J. Holden, "Fixed points and two-cycles of the discrete logarithm", pp. 405–415 in *Algorithmic number theory* (Sydney, 2002), edited by C. Fieker and D. R. Kohel, Lecture Notes in Comput. Sci. **2369**, Springer, Berlin, 2002. "Addenda and corrigenda" at arXiv:math.NT/0208028. MR 2005h:11277 Zbl 1058.11073

[Holden and Moree 2004] J. Holden and P. Moree, "New conjectures and results for small cycles of the discrete logarithm", pp. 245–254 in *High primes and misdemeanours*, edited by A. van der Poorten and A. Stein, Fields Inst. Commun. **41**, Amer. Math. Soc., Providence, RI, 2004. MR 2005d:11005 Zbl 1100.11005

[Holden and Moree 2006] J. Holden and P. Moree, "Some heuristics and results for small cycles of the discrete logarithm", *Math. Comp.* **75**:253 (2006), 419–449. MR 2006i:11145 Zbl 1116.11004

[Lindle 2008] N. Lindle, *A statistical look at maps of the discrete logarithm*, senior thesis, Rose-Hulman Institute of Technology, 2008, available at http://www.csse.rose-hulman.edu/images/docs/theses/NathanLindle2008.pdf.

[Mace 2009] M. L. Mace, "Discrete logarithm over composite moduli", REU technical report, Rose-Hulman Institute of Technology, 2009, available at http://www.rose-hulman.edu/~holden/REU/Reports/mace.pdf.

[Mishna 2004] M. Mishna, "How to use attribute grammars with ease and pleasure", 2004, available at http://www.math.sfu.ca/~mmishna/Publications/cook2.ps.

[Pollard 1978] J. M. Pollard, "Monte Carlo methods for index computation (mod *p*)", *Math. Comp.* **32**:143 (1978), 918–924. MR 58 #10684 Zbl 0382.10001

[Rivest et al. 1978] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", *Comm. ACM* **21**:2 (1978), 120–126. MR 83m:94003 Zbl 0368.94005

[Salvy 1997] B. Salvy, "Pollard's rho algorithm", worksheet, 1997, available at http://algo.inria.fr/libraries/autocomb/pollard-html/pollard1.html.

[Shepp and Lloyd 1966] L. A. Shepp and S. P. Lloyd, "Ordered cycle lengths in a random permutation", *Trans. Amer. Math. Soc.* **121** (1966), 340–357. MR 33 #3320 Zbl 0156.18705

Received: 2009-10-17 Revised: Accepted: 2010-06-19

Daniel.R.Cloutier@alumni.rose-hulman.edu
Rose–Hulman Institute of Technology,
Terre Haute, IN 47803, United States

holden@rose-hulman.edu
Rose–Hulman Institute of Technology,
Department of Mathematics, CM #125, 5500 Wabash Ave.,
Terre Haute, IN 47803, United States
http://www.rose-hulman.edu/~holden

Linear dependency for the difference in exponential regression

Indika Sathish and Diawara Norou

(Communicated by Kenneth S. Berenhaut)

In the field of reliability, a lot has been written on the analysis of phenomena that are related. Estimation of the difference of two population means have been mostly formulated under the no-correlation assumption. However, in many situations, there is a correlation involved. This paper addresses this issue. A sequential estimation method for linearly related lifetime distributions is presented. Estimations for the scale parameters of the exponential distribution are given under square error loss using a sequential prediction method. Optimal stopping rules are discussed using concepts of mean criteria, and numerical results are presented.

1. Introduction

In recent years, there has been a great deal of interest in looking at parameters and characterization of linearly related lifetime distributions, and more specifically of exponential types distributions. In the literature, estimation of the parameters using the sequential prediction method can be found in many areas such as statistical sciences, industrial quality control, communication science, computer simulations, genetics and many more. The sequential analysis method is carried out to determine improvements on the estimators and reduce noises related to the lifetime distributions. However, in many cases, when a pair of distributions are considered, the assumption of independence is assumed. There are contexts in which the assumption of independence is not realistic, such as in [Carpenter et al. 2006]. This paper extends the results that are proposed by including a correlation in estimating the difference parameter between two exponentially distributed functions. It is organized as follows. In Section 2, we present the basic results and the problem of interest. In Section 3, we present the sequential analysis method for the estimation of the difference of the scale parameters. Many works, such as [Mukhopadhyay and Hamdy 1984], have addressed the estimation of the difference

MSC2000: primary 60-00; secondary 60-08.

Keywords: survival, stopping rule, bivariate exponential.

of the location parameters of two distributions. Lai [2001] gives a thorough review of the sequential analysis technique along with challenges. The sections that follow are about the stopping rule technique and simulations.

2. Preliminaries and problem of interest

We consider the class of exponential family type probability distributions on the real line from [McCullagh and Nelder 1983]. The class is defined by the family of densities $\mathcal{G}$ with respect to the Lebesgue measure as

$$f(x;\theta,\varphi)=\exp\left\{\frac{\theta T(x)-b(\theta)}{a(\varphi)}+c(x,\varphi)\right\}, \tag{1}$$

where

- $f\in\mathcal{G}$,
- φ is a constant scale parameter, typically called the nuisance parameter,
- θ is a location parameter,
- $a(\varphi)$ and $c(x,\varphi)$ are specific functions of the scale parameter, and
- $b(\theta)$ and $T(x)$ are functions of the location parameter and variable x, respectively.

In fact, this exponential family density in (1) is a reformulation of the form given in [McCullagh and Nelder 1983] as they simplify $T(x)$ in (1) to simply x. Also, the expression (1) generalizes the exponential family type of distributions as described in [Terbeche et al. 2005] in the sense that

- if φ is known, (1) is the linear exponential family with canonical parameter θ;
- if φ is unknown, (1) may be used as a 2-parameter exponential family type.

As described in [McCullagh and Nelder 1983], this family includes the normal, exponential, gamma, and Poisson types of distributions. In this setting,

$$U=U(\theta)=\frac{\partial\log L(\theta,x)}{\partial\theta}=\frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)} \tag{2}$$

is the score function. Note that

- $E(U)=0$,
- $\operatorname{Var}(U)=E(U^2)=-E(\partial U/\partial\theta)=I(\theta),$ known as Fisher's information.

In the exponential family case, as in (1),

$$l(\theta,\varphi,x)=\log L(\theta,\varphi,x)=\frac{\theta T(x)-b(\theta)}{a(\varphi)}+c(x,\varphi),$$

$$U=\frac{\partial l}{\partial\theta}=\frac{T(x)-\partial b(\theta)/\partial\theta}{a(\varphi)},$$

$$E(U)=0 \implies E\big(T(x)\big)=\frac{\partial b(\theta)}{\partial\theta}=b'(\theta).$$

Based on some index set I, we now consider two classes of exponential families of random variables called $\mathbf{X}=(X_i)_{i\in I}$ and $\mathbf{Y}=(Y_i)_{i\in I}$, with densities

$$f(x_i;\theta,\varphi)=\exp\left\{\frac{\theta T(x_i)-b(\theta)}{a(\varphi)}+c(x_i,\varphi)\right\},\tag{3}$$

$$f(y_i;\tilde{\theta},\tilde{\varphi})=\exp\left\{\frac{\tilde{\theta}T(y_i)-\tilde{b}(\tilde{\theta})}{\tilde{a}(\tilde{\varphi})}+\tilde{c}(y_i,\tilde{\varphi})\right\}.\tag{4}$$

in the classes $\mathscr{G}_X$ and $\mathscr{G}_Y$, with the linear relationship

$$Y_i=aX_i+Z_i,\tag{5}$$

where $i\in I$, a is a fixed positive constant, and the Z_i are unknown random variables whose means are of interest.

The set I is an index countable set that could be finite or infinite. The linear relation described in (5) of association of random variables is not new, but is still a challenging problem. In fact, many authors [Carpenter et al. 2006; Iyer et al. 2002; 2004] have suggested its importance in applications.

Our goal is to estimate the parameter

$$\lambda=E_{\tilde{\theta}}[T(\mathbf{Y})]-aE_{\theta}[T(\mathbf{X})]\ ,\tag{6}$$

with square error loss. When $a=1$, this equation reduces to the difference between two dependent exponential family of distributions. The dependence concept is the innovation here as in many cases independence is assumed, even if it is known that there is great cost associated with that independence assumption.

3. Sequential analysis

We use the sequential estimation procedure to estimate the mean of the difference of two exponential families distributions with conjugate priors of the gamma or Bernoulli or Poisson types. This procedure helps address the problem in the small sample size case, maintaining a high power . The approach we use is Bayesian and we assume that $\pi_1(\theta)$ and $\pi_2(\tilde{\theta})$ are the conjugate priors given by

$$\pi_1(\theta)\propto\exp[t(\mu_1\theta-\varphi(\theta))],\quad \pi_2(\tilde{\theta})\propto\exp[s(\mu_2\tilde{\theta}-\tilde{\varphi}(\tilde{\theta}))].$$

This is not a new idea; Diaconis and Ylvisaker [1979] adopted this alternative to the maximum likelihood estimation regarding the parameter θ as a random variable with prior distribution, and the inference was based on the posterior distribution. They used this setting in the exponential family with conjugate prior distribution of the parameter θ given as

$$\pi(\theta) = \frac{\exp\{t(\mu\theta - \phi(\theta))\}}{\int \exp\{t(\mu\theta - \phi(\theta))\}d\theta}, \tag{7}$$

where $\theta \in \Theta$, t can be thought as prior sample size, and μ is the mean parameter. See also [Annis 2007].

In that regard, we see that $\mu_1 = E_{\pi_1}[\varphi'(\theta)]$ and $\mu_2 = E_{\pi_2}[\tilde{\varphi}'(\tilde{\theta})]$ are prior estimators of $E_\theta[T(\mathbf{X})]$ and $E_{\tilde{\theta}}[T(\mathbf{Y})]$, respectively.

Hence, following an idea from [Terbeche et al. 2005], the Bayes estimate of λ, based on a random sample of size n of $X_1, X_2, \dots, X_n$ of $\mathbf{X}$, and $Y_1, Y_2, \dots, Y_n$ of $\mathbf{Y}$ is given by

$$\begin{aligned}\hat{\lambda} &= \hat{\lambda}(\mathbf{X}, \mathbf{Y}) = \hat{\lambda}(X_1, \dots, X_n, Y_1, \dots, Y_n)\\ &= E[\lambda | X_1, \dots, X_n, Y_1, \dots, Y_n]\\ &= E[\tilde{b}'(\tilde{\theta})|Y_1, \dots, Y_n] - aE[b'(\theta)|X_1, \dots, X_n],\end{aligned}$$

where

$$E[b'(\theta)|X_1, \dots, X_n] = \frac{n\bar{T}_n^{\mathbf{X}} + t\mu_1}{n+t}, \quad E[\tilde{b}'(\tilde{\theta})|Y_1, \dots, Y_n] = \frac{n\bar{T}_n^{\mathbf{Y}} + s\mu_2}{n+s}, \tag{8}$$

with

$$\bar{T}_n^{\mathbf{X}} = \frac{T(X_1) + \ldots + T(X_n)}{n}, \quad \bar{T}_n^{\mathbf{Y}} = \frac{T(Y_1) + \ldots + T(Y_n)}{n}.$$

Hence,

$$\hat{\lambda} = \frac{n\bar{T}_n^{\mathbf{Y}} + s\mu_2}{n+s} - a\frac{n\bar{T}_n^{\mathbf{X}} + t\mu_1}{n+t}. \tag{9}$$

The asymptotic estimate for the parameter as $n \longrightarrow \infty$ is

$$\hat{\lambda} = \bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}. \tag{10}$$

A criteria for stopping the estimation of λ is developed. When $t = s$,

$$\hat{\lambda} = \frac{n(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + t(\mu_2 - a\mu_1)}{n+t} = \frac{n}{n+t}(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + \frac{t}{n+t}(\mu_2 - a\mu_1).$$

When $t = s = n$,

$$\hat{\lambda} = \frac{(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + (\mu_2 - a\mu_1)}{2}. \tag{11}$$

In the sequential analysis idea, the sample size is not predetermined. Hence, a natural question to ask is when is the sample size large enough to make conclusions.

4. Stopping rules

The Bayes risk of the estimate $\hat{\lambda}$ of λ with respect to the prior $\pi(\theta)$ in (7) is

$$r(\theta, \hat{\lambda}) = E[R(\theta, \hat{\lambda})],$$

where $R(\theta, \hat{\lambda}) = E[L(\theta, \hat{\lambda})]$ and $L(\theta, \hat{\lambda}) = (\lambda - \hat{\lambda})^2$ is the loss function.

In this setting, the Bayes risk is given by

$$\begin{aligned} r(\pi_1, \pi_2) &= r(\hat{\lambda}(\mathbf{X}, \mathbf{Y})) \\ &= E_{(\mathbf{XY})}[E_{\lambda|(\mathbf{X},\mathbf{Y})}(\hat{\lambda}(\mathbf{X}, \mathbf{Y}) - \lambda)^2] \\ &= E_{(\mathbf{X},\mathbf{Y})}[\mathrm{Var}(\lambda|(\mathbf{X}, \mathbf{Y}))] \\ &= E_{(\mathbf{X},\mathbf{Y})}\big[\mathrm{Var}(\tilde{b}'(\tilde{\theta}) - ab'(\theta)|(\mathbf{X}, \mathbf{Y}))\big] \\ &= E_{(\mathbf{X},\mathbf{Y})}\big[\mathrm{Var}(\tilde{b}'(\tilde{\theta})) + a^2\,\mathrm{Var}(b'(\theta)) - 2a\rho\sqrt{\mathrm{Var}(\tilde{b}'(\tilde{\theta}))}\sqrt{\mathrm{Var}(b'(\theta))}\big], \end{aligned}$$

and the upper bound is achieved using the idea of Equation (4) in [Terbeche et al. 2005]. It is given by

$$r(\pi_1, \pi_2) = E_{\mathbf{Y}}\left[E_{\tilde{\theta}|\mathbf{Y}}\left|\frac{\tilde{b}''(\tilde{\theta})}{n+s}\right|\right] + a^2 E_{\mathbf{X}}\left[E_{\theta|\mathbf{X}}\left|\frac{b''(\theta)}{n+t}\right|\right], \tag{12}$$

with equality achieved in (12) when $\rho = \mathrm{corr}(\tilde{b}'(\tilde{\theta}), b'(\theta)) = \mathrm{corr}(\mathbf{X}, \mathbf{Y}) \geq 0$ is minimized.

Considering the loss function

$$L(\lambda, \hat{\lambda}, n) = (\lambda - \hat{\lambda})^2 + cn, \tag{13}$$

where c can be looked at as the cost of sampling, and the decision rule $\Delta = (\tau, \delta)$, where $\tau = \tau_n(\mathbf{x}, \mathbf{y})$ is the stopping rule and $\delta = \delta_n(\mathbf{x}, \mathbf{y})$ is the decision rule, we have that the Bayes risk to minimize from a suitable sample size n obtained sequentially given by

$$\begin{aligned} r(\tau, \pi_1, \pi_2) &= E_{(\mathbf{X},\mathbf{Y},\tau)}\left[\frac{U_n}{n+t} + \frac{V_n}{n+s} - 2a\rho\sqrt{[\mathrm{Var}(\tilde{b}'(\tilde{\theta}))]}\sqrt{[\mathrm{Var}(b'(\theta))]} + cn\right] \\ &= E_{(\mathbf{Y},\tau)}\left[\frac{U_n}{n+t}\right] + E_{(\mathbf{X},\tau)}\left[\frac{V_n}{n+s}\right] \\ &\qquad + E_{(\mathbf{X},\mathbf{Y},\tau)}\big[-2a\rho\sqrt{[\mathrm{Var}(\tilde{b}'(\tilde{\theta}))]}\sqrt{[\mathrm{Var}(b'(\theta))]} + cn\big], \end{aligned}$$

where $U_n = E_{\mathbf{Y},\tau}|\tilde{b}''(\tilde{\theta})|$ and $V_n = E_{\mathbf{X},\tau}|b''(\theta)|$.

Using ideas in [Terbeche et al. 2005] to achieve the upper bound in (12), the stopping rule criteria can be expressed as if

$$U_n \leq c(n+t)^2 \quad \text{or} \quad V_n \leq c(n+s)^2,$$

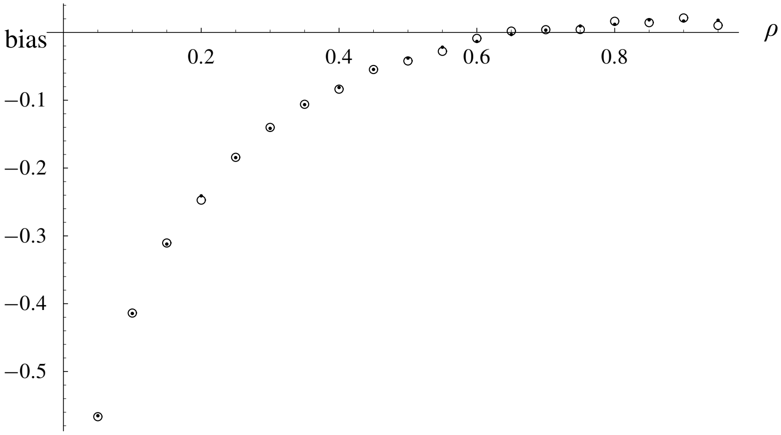


Figure 1. Graph of the bias from ρ for $c = 0$.

then take another pair of observations. Otherwise, stop the collection process. That is the estimation of the difference of the two exponential distributions can be evaluated from the available informative sample. In other words, the stopping variable is defined by the quantity

$$n \geq \min\left\{\sqrt{\frac{U_n}{c}} - t, \sqrt{\frac{V_n}{c}} - s\right\}. \tag{14}$$

In order to study the optimized stopping rule in (14) and its efficiency, a numerical simulation technique is provided in Section 5. We consider two exponentially related distributions with gamma priors.

5. Simulation

We have described a methodology to compare the mean difference between two exponential distributions that are linearly related. In this section, we show an example of a simulation data of the related bivariate exponential distribution with the different values of the correlations ρ.

Since we consider two dependent random variables, we create one exponential random variable and create the other one with the desired correlation ρ. We generate sample data of size 50. We assume a coefficient of linear relationship $a = 1$ of simultaneous occurrence as described in [Marshall and Olkin 1967], and $c = 0$ and $c = 0.25$ in (13) over 5000 runs. The simulation was carried out using SAS.

The results of the two figures show that data does not need to be large to achieve convergence. The pattern is the same regardless of the number of runs. Figures 1 and 2 give the bias of the mean difference for $c = 0$ and $c = 0.25$, respectively. The convergence is justified by the maximal error we allowed to reach based on the stopping rule, when the data generation and bias are computed at three and five

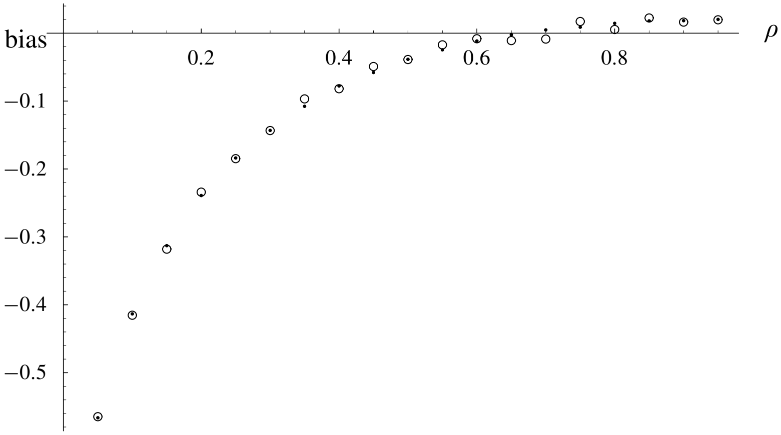


Figure 2. Graph of the bias from ρ for $c = 0.25$.

decimal places (circles and dots, respectively). The algorithm performs very well even when the sample size is small, showing great robustness.

The resulting plot of the bias is very helpful in explaining the effectiveness of the estimator. When the correlation is present, this new estimator should be considered. Furthermore, the choice of the cost of resampling c does not affect significantly in the error estimation. Setting $c = 0.25$ as in Figure 2 shows the same trend as for Figure 1. The risk is then minimized considerably when the correlation is significant.

6. Conclusion

The proposed sequential parametric procedure in the estimation of the difference of two exponential distribution is quite useful and relevant. This sequential estimation for the bivariate distributions of the exponential type families is used to get an estimate of the mean difference. It is more efficient in terms of bias.

Acknowledgements

The authors are grateful to the referees for their detailed suggestions, comments and insights, which improved the quality of the paper considerably.

References

[Annis 2007] D. H. Annis, "A note on quasi-likelihood for exponential families", *Statist. Probab. Lett.* **77**:4 (2007), 431–437. MR 2339048 Zbl 1108.62022

[Carpenter et al. 2006] M. Carpenter, N. Diawara, and Y. Han, "A new class of bivariate survival and reliability models", *Amer. J. Math. Management Sci.* **26**:1-2 (2006), 163–184. MR 2007m:62144 Zbl 1154.62337

[Diaconis and Ylvisaker 1979] P. Diaconis and D. Ylvisaker, "Conjugate priors for exponential families", *Ann. Statist.* **7**:2 (1979), 269–281. MR 80f:62016 Zbl 0405.62011

[Iyer and Manjunath 2004] S. K. Iyer and D. Manjunath, "Correlated bivariate sequences for queueing and reliability applications", *Comm. Statist. Theory Methods* **33**:2 (2004), 331–350. MR 2045319 Zbl 1066.62054

[Iyer et al. 2002] S. K. Iyer, D. Manjunath, and R. Manivasakan, "Bivariate exponential distributions using linear structures", *Sankhyā Ser. A* **64**:1 (2002), 156–166. MR 1968380

[Lai 2001] T. L. Lai, "Sequential analysis: some classical problems and new challenges", *Statist. Sinica* **11**:2 (2001), 303–408. MR 2002d:62001 Zbl 1037.62081

[Marshall and Olkin 1967] A. W. Marshall and I. Olkin, "A multivariate exponential distribution", *J. Amer. Statist. Assoc.* **62** (1967), 30–44. MR 35 #6241 Zbl 0147.38106

[McCullagh and Nelder 1983] P. McCullagh and J. A. Nelder, *Generalized linear models*, Monographs on Statistics and Applied Probability **37**, Chapman & Hall, London, 1983. MR 85k:62161 Zbl 0588.62104

[Mukhopadhyay and Hamdy 1984] N. Mukhopadhyay and H. I. Hamdy, "On estimating the difference of location parameters of two negative exponential distributions", *Canad. J. Statist.* **12**:1 (1984), 67–76. MR 86b:62139 Zbl 0543.62061

[Terbeche et al. 2005] M. Terbeche, B. O. Oluyede, and A. Barbour, "On sequential and fixed designs for estimation with comparisons and applications", *SORT* **29**:2 (2005), 217–233. MR 2208558 Zbl 05633857

Received: 2009-11-09 Revised: 2010-03-26 Accepted: 2010-04-18

sindika@odu.edu
Old Dominion University,
Department of Mathematics and Statistics,
4700 Elkhorn Avenue, Norfolk, VA 23529, United States

ndiawara@odu.edu
Old Dominion University,
Department of Mathematics and Statistics,
4700 Elkhorn Avenue, Norfolk, VA 23529, United States
http://www.odu.edu/~ndiawara

The probability of relatively prime polynomials in $\mathbb{Z}_{p^k}[x]$

Thomas R. Hagedorn and Jeffrey Hatley

(Communicated by Arthur T. Benjamin)

Let $\mathrm{P}_R(m, n)$ denote the probability that two randomly chosen monic polynomials $f, g \in R[x]$ of degrees m and n, respectively, are relatively prime. Let $q = p^k$ be a prime power. We establish an explicit formula for $\mathrm{P}_R(m, 2)$ when $R = \mathbb{Z}_q$, the ring of integers mod q.

1. Introduction

Given two polynomials $f(x)$, $g(x)$ chosen at random, what is the probability that they are relatively prime? For a ring R, we say that two polynomials $f, g \in R[x]$ are relatively prime if there is no monic polynomial of positive degree that divides both f and g. Let $\mathrm{P}_R(m, n)$ denote the probability that two randomly chosen monic polynomials f, $g \in R[x]$ of degrees m and n, respectively, are relatively prime. If R has an infinite number of elements, then $\mathrm{P}_R(m, n) = 1$, so we restrict our attention to finite rings R. Let $R = \mathbb{F}_q$, the finite field with q elements. The formula, $\mathrm{P}_{\mathbb{F}_q}(m, m) = 1 - 1/q$ was proved in [Corteel et al. 1998]. When $q = p = 2$, Reifegerste [2000] gave a combinatorial proof that $\mathrm{P}_{\mathbb{F}_2}(m, m) = 1/2$. Benjamin and Bennett subsequently found a beautifully simple proof generalizing these results:

Theorem 1.1 [Benjamin and Bennett 2007]. *If $m, n \geq 1$, then $\mathrm{P}_{\mathbb{F}_q}(m, n) = 1 - \dfrac{1}{q}$.*

This can be generalized in at least two ways. Hou and Mullen [2009] have generalized Theorem 1.1 by considering the problem of relatively prime polynomials in several variables over a finite field. In earlier work, Gao and Panario [2006] considered the probability distribution of the greatest common divisor of l randomly chosen monic single-variable polynomials in $\mathbb{F}_q[x]$ with degrees $n_1, \ldots, n_l$ as the $n_i \to \infty$. In this paper, we restrict ourselves to single-variable polynomials and explore a different perspective.

MSC2000: 11C20, 13B25, 13F20.

Keywords: relatively prime polynomials.

The authors would like to thank The College of New Jersey for its support of undergraduate research.

As the formula in Theorem 1.1 only depends on the number of elements in the field $\mathbb{F}_q$, one can ask whether the same formula holds when R is another ring with q elements. For example, if $R = \mathbb{Z}_q$, the integers mod q, does the same formula hold? It does not, but the formula for $\mathrm{P}_{\mathbb{F}_q}(m, n)$ can be viewed as a first approximation to the formula for $\mathrm{P}_{\mathbb{Z}_q}(m, n)$. In this paper, we prove an explicit formula for $\mathrm{P}_{\mathbb{Z}_{p^k}}(m, 2)$ for p odd.

For each positive integer k, we define a monic polynomial $f_k(x) \in \frac{1}{2}\mathbb{Z}[x]$ by

$$f_k(x) = x^{2k} + (1-x)\sum_{i=0}^{(k-3)/2} x^{(k+3)/2+3i} + \frac{1}{2}\sum_{i=0}^{k-1}(-x)^i + \tfrac{1}{2}x^{(k-1)/2} - 1,$$

for k odd, and

$$f_k(x) = x^{2k} + (1-x)\sum_{i=1}^{k/2-1} x^{2k-3i} - \frac{1}{2}\sum_{i=1}^{k-1}(-x)^i - x^{k/2+1} + \tfrac{3}{2}x^{k/2} - 1,$$

for k even. The polynomial $f_k(x)$ has degree $2k$ and its coefficients have absolute value at most 2.

Theorem 1.2. *Let p be an odd prime and let $m, k \geq 1$ be integers. The probability that two randomly chosen monic polynomials in $\mathbb{Z}_{p^k}[x]$ of degrees m and 2, respectively, are relatively prime is*

$$\mathrm{P}_{\mathbb{Z}_{p^k}}(m, 2) = 1 - \frac{1}{p^{3k}} f_k(p).$$

When $k = 1$, we rediscover $\mathrm{P}_{\mathbb{F}_p}(m, 2) = 1 - 1/p$. For small values of k, we have

$$\begin{aligned}
\mathrm{P}_{\mathbb{Z}_{p^2}}(m, 2) &= 1 - \frac{1}{p^2} + \frac{1}{p^4} - \frac{2}{p^5} + \frac{1}{p^6},\\
\mathrm{P}_{\mathbb{Z}_{p^3}}(m, 2) &= 1 - \frac{1}{p^3} + \frac{1}{p^5} - \frac{1}{p^6} - \frac{1}{2p^7} + \frac{1}{2p^9},\\
\mathrm{P}_{\mathbb{Z}_{p^4}}(m, 2) &= 1 - \frac{1}{p^4} + \frac{1}{p^6} - \frac{1}{p^7} + \frac{1}{2p^9} - \frac{1}{p^{10}} - \frac{1}{2p^{11}} + \frac{1}{p^{12}}.
\end{aligned}$$

As an immediate corollary to Theorem 1.2, we obtain:

Corollary 1.3. *Given $k \geq 1$, there exists a monic polynomial*

$$g_k(x) = \sum a_i x^i \in \tfrac{1}{2}\mathbb{Z}[x]$$

with degree $2k-2$ and $|a_i| \leq 2$, such that

$$\mathrm{P}_{\mathbb{Z}_{p^k}}(m, 2) = 1 - \frac{1}{p^k} + \frac{1}{p^{3k}} g_k(p) \quad \textit{for all odd primes } p \textit{ and all } m \geq 1.$$

We obtain Theorem 1.2 and its corollary by adapting the arguments of Benjamin and Bennett [2007], who proved Theorem 1.1 by a clever use of the Euclidean algorithm in $\mathbb{F}_q[x]$. While $\mathbb{Z}_{p^k}[x]$ does not have the Euclidean algorithm, due to the existence of noninvertible elements in $\mathbb{Z}_{p^k}$, it does have a division algorithm for monic polynomials. This division algorithm, together with some facts about polynomial factorization of quadratics in $\mathbb{Z}_{p^k}[x]$, suffices to prove Theorem 1.2 for odd primes p. It appears that our arguments can also be used to prove the formula for $\mathrm{P}_{\mathbb{Z}_{p^k}}(m, 2)$ when $p = 2$, and also a formula for $\mathrm{P}_{\mathbb{Z}_{p^k}}(m, 3)$, but the details are much more involved and have not yet been fully worked through. However, the present approach does not seem able to establish a formula for $\mathrm{P}_{\mathbb{Z}_{p^k}}(m, n)$ for general $m, n \geq 4$ as the number of cases to consider in the proof grows as a function of $\min(m, n)$.

2. Arithmetic in $\mathbb{Z}_{p^k}[x]$

In this section, we establish some basic results on the rings $\mathbb{Z}_{p^k}$ and $\mathbb{Z}_{p^k}[x]$. Recall that $\mathbb{Z}_n$ denotes the ring of integers mod n. We will make use of Hensel's lemma [Gouvêa 1997, page 70] in the following form:

Lemma 2.1 (Hensel's lemma). *Let $f(x) \in \mathbb{Z}_{p^k}[x]$ be a polynomial and denote its reduction mod p by $\bar{f}(x) \in \mathbb{Z}_p[x]$. Suppose there exists $u_0 \in \mathbb{Z}_p$ with $\bar{f}(u_0) = 0$ in $\mathbb{Z}_p$ and $\bar{f}'(u_0) \neq 0$ in $\mathbb{Z}_p$. Then there exists a unique $u \in \mathbb{Z}_{p^k}$, with $f(u) = 0$ in $\mathbb{Z}_{p^k}$ and $u \equiv u_0 \bmod p$.*

We start by counting the squares in $\mathbb{Z}_{p^k}$ and its unit subgroup $\mathbb{Z}^*_{p^k}$.

Lemma 2.2. *Let p be an odd prime and $k \geq 1$.*

(a) *$\mathbb{Z}^*_{p^k}$ has $\frac{1}{2}p^{k-1}(p-1)$ squares.*

(b) *Let d be even, with $0 \leq d < k$. There are $\frac{1}{2}(p-1)p^{k-1-d}$ nonzero squares $x \in \mathbb{Z}_{p^k}$ with $x \in p^d\mathbb{Z}_{p^k} \setminus p^{d+1}\mathbb{Z}_{p^k}$.*

(c) *There are $1 + \dfrac{1}{2(p+1)}(p^{k+1} - p^{1-k+2[k/2]})$ squares in $\mathbb{Z}_{p^k}$.*

Proof. (a) We first note that the $(p-1)/2$ squares $x = 1^2, \dots, (\frac{p-1}{2})^2$ are distinct nonzero squares in both $\mathbb{Z}_p$ and $\mathbb{Z}_{p^k}$. Now consider a unit $u \in \mathbb{Z}_{p^k}$ satisfying $u \equiv 1 \bmod p$. Letting $f(x) = x^2 - u \in \mathbb{Z}_{p^k}[x]$, and $u_0 = 1$, by Lemma 2.1, u is a square in $\mathbb{Z}_{p^k}$. Thus the p^{k-1} units $u \in \mathbb{Z}_{p^k}$ with $u \equiv 1 \bmod p$ are squares. Hence, the $\frac{1}{2}p^{k-1}(p-1)$ distinct units xu are all squares and every unit square can be seen to be of this form.

(b) Let $x \in \mathbb{Z}_{p^k}$ satisfy $x \in p^d\mathbb{Z}_{p^k} \setminus p^{d+1}\mathbb{Z}_{p^k}$. Let $x = (p^t u)^2 = p^{2t}u^2$, where u is a unit. To satisfy the given conditions, $t = d/2$, u^2 is a unit square in $\mathbb{Z}^*_{p^k}$, and $u^2 \equiv u_1^2 \bmod p^{k-d}$. Hence, the number of distinct x equals the number of unit squares in $\mathbb{Z}_{p^{k-d}}$, which is given by (a).

(c) Every nonzero square can be written as $p^{2d}u$, where u is a unit square and $0 \le 2d < n$. Counting the square 0, the total sum is, thanks to (b),

$$1+\tfrac{1}{2}(p-1)\sum_{d=0}^{[(k-1)/2]} p^{k-1-2d}.$$

This expression simplifies to the claimed formula. □

For $g(x) = x^2 + bx + c \in \mathbb{Z}_{p^k}[x]$, define the discriminant $\Delta_g = b^2 - 4c$. As when $k = 1$, we can describe the number of roots of $g(x) \in \mathbb{Z}_{p^k}[x]$ using Δ_g.

Lemma 2.3. *Let p be an odd prime, $k \ge 1$, and $g(x) = x^2 + bx + c \in \mathbb{Z}_{p^k}[x]$.*

(a) *Δ is a square mod p^k if and only if g is reducible.*

(b) *If $\Delta \equiv 0 \bmod p^k$, then g has the $p^{[k/2]}$ roots given by $\dfrac{-b}{2}+p^{[(k+1)/2]}t \bmod p^k$, where $t = 1, \dots, p^{[k/2]}$.*

(c) *Suppose $\Delta \equiv p^d u \bmod p^k$ is a nonzero square with $0 \le d < k$, d even, $u \in \mathbb{Z}^*_{p^k}$ a square. Choose a such that $u \equiv a^2 \bmod p^k$. Then g has the $2p^{d/2}$ roots*

$$-\tfrac{1}{2}b \pm \tfrac{1}{2}ap^{d/2} + tp^{k-d/2} \mod p^k, \qquad \text{where } t = 1, \dots, p^{d/2}.$$

Proof. Since p is odd, we have $g(x) = (x+b/2)^2 - \Delta/4$. Hence $r = -(b+z)/2$ is a root of $g(x)$ if and only if z is a solution of the equation $z^2 \equiv \Delta \bmod p^k$. Condition (a) is thus proved. Condition (b) follows as well as the roots of the equation $z^2 \equiv 0 \bmod p^k$ are $z \equiv p^{[(k+1)/2]}t \bmod p^k$, for $t = 1, \dots, p^{[k/2]}$, or equivalently, $z \equiv 2p^{[(k+1)/2]}t \bmod p^k$, for $t = 1, \dots, p^{[k/2]}$. (c) By the hypothesis, d is even and $a \not\equiv 0 \bmod p$. The solutions to the equation $z^2 \equiv p^d a^2 \bmod p^k$ have the form $z \equiv p^{d/2}w \bmod p^k$, where $w \in \mathbb{Z}_{p^k}$ is a solution of $x^2 \equiv a^2 \bmod p^{k-d}$. Hensel's lemma (using the polynomial $f(x) = x^2 - a^2$), shows that the solutions to this latter equation are the $w \in \mathbb{Z}_{p^k}$ satisfying $w \equiv \pm a \bmod p^{k-d}$. Thus $w = \pm a + tp^{k-d}$, for $t = 1, \dots, p^d$, or equivalently, as 2 is a unit mod p^d, $w = \pm a + 2tp^{k-d}$ for $t = 1, \dots, p^d$. Now two roots $z = p^{d/2}w$ and $z_1 = p^{d/2}w_1$ are equal precisely when the signs in the expressions for w and w_1 agree and the respective parameters t and t_1 satisfy $t \equiv t_1 \bmod p^{d/2}$. Hence we have shown that the original equation $z^2 \equiv p^d a^2 \bmod p^k$ has the $2p^{d/2}$ distinct roots given by $z = \pm ap^{d/2} + 2tp^{k-d/2}$, for $t = 1, \dots, p^{d/2}$. □

Lemma 2.4. *Let p be an odd prime and $k \ge 1$.*

(a) *Given $\Delta \in \mathbb{Z}_{p^k}$, there are p^k monic, quadratic polynomials $g \in \mathbb{Z}_{p^k}[x]$ with $\Delta_g \equiv \Delta \bmod p^k$.*

(b) *There are*
$$\frac{p^k}{2(p+1)}(p^{k+1}+2p^k-p-p^{k-2[k/2]}-1)$$
monic, irreducible, quadratic polynomials $g \in \mathbb{Z}_{p^k}[x]$.

Proof. If $g = x^2+bx+c$, then $\Delta_g = b^2-4c$. Since 4 is invertible mod p^k, for every Δ, $b \in \mathbb{Z}_{p^k}$, there is a unique choice of c such that $\Delta_g \equiv \Delta \bmod p^k$. Since there are p^k choices for b, (a) is proved. Now g is irreducible precisely when Δ_g is not a square. Let S be the number of squares in $\mathbb{Z}_{p^k}$. Then for each $b \in \mathbb{Z}_{p^k}$, there are $p^k - S$ choices for c such that b^2-4c is not a square. Thus, using the formula for S given by Lemma 2.2(c), there are
$$p^k(p^k-S) = \frac{p^k}{2(p+1)}\big(p^{k+1}+2p^k-2p+p^{1-k+2[k/2]}-2\big)$$
irreducible polynomials g. Simplification gives (b). □

Given a monic, quadratic polynomial $g \in \mathbb{Z}_{p^k}[x]$, we define the set
$$A_g = \{h \in \mathbb{Z}_{p^k}[x] : \deg h \le 1 \text{ and } g, h \text{ are not relatively prime}\},$$
and let $|A_g|$ denote its cardinality. We note that in the definition of A_g, we allow nonmonic polynomials h.

Lemma 2.5. *Let p be an odd prime and $g(x)$ be a monic quadratic polynomial in* $\mathbb{Z}_{p^k}[x]$.

(a) *If* $\Delta_g \equiv 0 \bmod p^k$, *then*
$$|A_g| = p^{k-[k/2]}\left(\frac{p^{2[k/2]+1}+1}{p+1}\right).$$

(b) *Assume* $\Delta_g \in \mathbb{Z}_{p^k}$ *is a nonzero square. Let* $\Delta_g \equiv p^d v \bmod p^k$, *where d is even,* $0 \le d < k$, *and* $v \in (\mathbb{Z}^*_{p^k})^2$. *Then*
$$|A_g| = 2p^{k-d/2}\left(\frac{p^{d+1}+1}{p+1}\right) - p^{d/2}.$$

Proof. We first note that a linear factor of $g(x)$ must have the form $u(x-r)$, where $u, r \in \mathbb{Z}_{p^k}$, u is a unit, and r is a root of g. Therefore, the elements $h(x) \in A_g$ are exactly the polynomials $h(x) = \alpha(x-r)$, for some $\alpha \in \mathbb{Z}_{p^k}$ and some root $r \in \mathbb{Z}_{p^k}$ of g. Hence, to calculate $|A_g|$, we need to count the number of distinct $h(x)$ of this form.

Suppose r_1 and r_2 are two roots of g and $\alpha(x-r_1) \equiv \beta(x-r_2) \bmod p^k$. Then $\beta \equiv \alpha \bmod p^k$ and $\alpha(r_1-r_2) \equiv 0 \bmod p^k$. Let $\alpha = p^s u$, with $u \in \mathbb{Z}^*_{p^k}$. If $s = k$, then $\alpha = 0$ is the only choice. Now suppose $s < k$. Then there are $p^{k-s-1}(p-1)$ distinct choices for u giving rise to distinct α. For each such α, we need to calculate the number of roots of g in $\mathbb{Z}_{p^{k-s}}$. To proceed further, we need to have a description of the roots.

Writing $g(x)=x^2+bx+c$, in case (a), the roots of g are $r=-b/2+p^{[(k+1)/2]}t$, for $t=1,\dots,p^{[k/2]}$ by Lemma 2.3. If $[k/2]\le s<k$, for each choice of $\alpha=p^s u$, there is exactly one factor $\alpha(x-r) \bmod p^k$. As there are $p^{k-s-1}(p-1)$ choices for u, and hence α, we obtain the same number of distinct factors $\alpha(x-r)$ for each s. If $0\le s\le[k/2]$, then for each choice of $\alpha=p^s u$, there are $p^{[k/2]-s}$ distinct factors $\alpha(x-r) \bmod p^k$. Hence there are $p^{k+[k/2]-2s-1}(p-1)$ distinct factors $\alpha(x-r) \bmod p^k$ for each s. In total then, we have

$$\begin{aligned}|A_g| &= \sum_{s=0}^{[k/2]}(p-1)p^{k+[k/2]-2s-1}+\Bigl(\sum_{s=[k/2]+1}^{k-1}(p-1)p^{k-s-1}+1\Bigr)\\ &= \sum_{s=0}^{[k/2]}(p-1)p^{k+[k/2]-2s-1}+p^{k-[k/2]-1}=p^{k-[k/2]}\Bigl(\frac{p^{2[k/2]+1}+1}{p+1}\Bigr),\end{aligned}$$

where the last equality is obtained by evaluating a geometric sum. We thus obtain the desired formula for case (a). In case (b), by Lemma 2.3, the roots of g are $-\frac{1}{2}b\pm\frac{1}{2}ap^{d/2}+tp^{k-d/2} \bmod p^k$, where $a^2\equiv v \bmod p^k$, $t=1,\dots,p^{d/2}$. As in case (a), we let $\alpha=p^s u$, and consider the number of distinct factors $h(x)=\alpha(x-r)$ for each choice of s. When $s=k$, $h(x)=\alpha=0$ is the only factor. There are three additional cases:

(1) Suppose $k>s\ge k-d/2$. Then $k-s\le d/2$ and all the roots of g are equivalent mod p^{k-s}. Since there are $p^{k-s-1}(p-1)$ distinct choices for α, there are the same number of distinct factors $\alpha(x-r)$.

(2) Suppose $k-d/2>s\ge d/2$. Then $d/2<k-s\le k-d/2$ and the roots of g determine two equivalence classes mod p^{k-s}. Thus for each s, there are a total of $2p^{k-s-1}(p-1)$ distinct factors $\alpha(x-r)$.

(3) Suppose $d/2\ge s\ge 0$. Then the roots of g determine $2p^{d/2-s}$ equivalence classes mod p^{k-s} for each α. Thus there are a total of $2p^{k+d/2-2s-1}(p-1)$ distinct factors $\alpha(x-r)$, for each s.

In total, when $d<k-1$, we have for $|A_g|$ the value

$$\begin{aligned}&\sum_{s=0}^{d/2}2(p-1)p^{k+d/2-2s-1}+\Bigl(\sum_{s=d/2+1}^{k-d/2-1}2(p-1)p^{k-s-1}+\sum_{s=k-d/2}^{k-1}(p-1)p^{k-s-1}+1\Bigr)\\ &\qquad=\sum_{s=0}^{d/2}2(p-1)p^{k+d/2-2s-1}+2p^{k-d/2-1}-p^{d/2},\end{aligned}$$

which simplifies to the formula stated in (b). When $d=k-1$, the second summation does not appear, and

$$\begin{aligned}|A_g| &= \sum_{s=0}^{d/2} 2(p-1)p^{k+d/2-2s-1} + \left(\sum_{s=k-d/2}^{k-1} (p-1)p^{k-s-1} + 1 \right) \\ &= \sum_{s=0}^{d/2} 2(p-1)p^{k+d/2-2s-1} + p^{d/2},\end{aligned}$$

which again simplifies to the stated formula for (b). □

3. Proof of the main theorem

In this section, we let $q = p^k$. To prove Theorem 1.2, we will count the number of polynomial pairs (f, g), where $f, g \in \mathbb{Z}_q[x]$ are not relatively prime. Let $f(x)$, $g(x)$ be monic polynomials. Then by the division algorithm, there is a unique choice of polynomials $q(x), r(x) \in \mathbb{Z}_q[x]$, with $q(x)$ monic, satisfying

$$f(x) = g(x)q(x) + r(x), \tag{1}$$

where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. Thus the pair (f, g) is uniquely determined by the triple $(g, q(x), r(x))$. From (1), any common divisor of f and g is a common divisor of g and r and vice-versa. We define

$$\begin{aligned}S_{m,d,q} &= \{(f, g) : f, g \in \mathbb{Z}_q[x] \text{ monic with } \deg\ f = m,\ \deg\ g = d, \\ &\qquad\qquad f \text{ and } g \text{ not relatively prime}\}, \\ T_{m,q} &= \{(g, r) : g, r \in \mathbb{Z}_q[x] \text{ with } g \text{ monic of degree } m,\ \deg\ r < m, \\ &\qquad\qquad g \text{ and } r \text{ not relatively prime}\}.\end{aligned}$$

Lemma 3.1. *If $m \geq d$, then $|S_{m,d,q}| = q^{m-d}|T_{d,q}|$.*

Proof. Let $(g, r) \in T_{d,q}$. Then each of the q^{m-d} monic polynomials $q(x)$ with degree $m-d$ gives rise via (1) to a unique pair $(f, g) \in S_{m,d,q}$. Conversely, the inverse map

$$(f, g) \mapsto (g, q, r) \mapsto (g, r)$$

is a q^{m-d}-to-1 map from $S_{m,d,q}$ to $T_{d,q}$. □

Thus, proving Theorem 1.2 is reduced to calculating $|T_{2,q}|$. We begin with:

Proposition 3.2. $|T_{1,q}| = q$.

Proof. If $(g, r) \in T_{1,q}$, then $g(x) = x - c$. For g and r to have a common factor, $r = 0$. Hence $T_{1,q}$ consists of the q pairs $(x - c, 0)$. □

We now determine $|T_{2,q}|$. By Lemma 2.3, we have $|T_{2,q}| = B_1 + B_2 + B_3$, where the B_i are defined by

$$B_1 = \bigl|\{(g,r) \in T_{2,q} : g \text{ is irreducible}\}\bigr|,$$

$$B_2 = \bigl|\{(g,r) \in T_{2,q} : \Delta_g \equiv 0 \bmod p^k\}\bigr|,$$

$$B_3 = \bigl|\{(g,r) \in T_{2,q} : \Delta_g \bmod p^k \text{ is a square, and, for each } d < k,\\ \Delta_g \equiv 0 \bmod p^d \text{ and } \Delta_g \not\equiv 0 \bmod p^{d+1}\}\bigr|.$$

Lemma 3.3. (a) $B_1 = \dfrac{p^k}{2(p+1)}(p^{k+1} + 2p^k - p - p^{k-2[k/2]} - 1)$.

(b) $B_2 = p^{2k-[k/2]}\left(\dfrac{p^{2[k/2]+1}+1}{p+1}\right)$.

(c) $B_3 = \dfrac{p^{2k-1-[(k-1)/2]} - p^{2k}}{2(p+1)(p^2+p+1)}\alpha$, *where*

$$\alpha = (p+1)(p^2+p+1) - 2p^{k+1}(p+1)^2 - 2p^{k-[(k-1)/2]}(p + p^{-[(k-1)/2]}).$$

Proof. (a) Assume $g \in \mathbb{Z}_{p^k}[x]$ is a monic, irreducible, quadratic polynomial. Since g has no factors, $(g,r) \in T_{2,q}$ only when $r = 0$. Hence, B_1 equals the number of monic, irreducible quadratic polynomials, which is given by Lemma 2.4.

(b) Assume $g \in \mathbb{Z}_{p^k}[x]$ is a monic quadratic with $\Delta_g \equiv 0 \bmod p^k$. By Lemma 2.4, there are p^k such g. For each g, $|A_g|$ is given by Lemma 2.5(a). Thus

$$B_2 = p^k|A_g|.$$

(c) If $(g,r) \in T_{2,q}$ is included in the pairs counted for B_3, then $\Delta_g = p^d u$, where $0 \le d < k$, d even, and $u \in \mathbb{Z}^*_{p^k}$ is a square. For a fixed d, u, satisfying these conditions, there are p^k polynomials g with $\Delta_g = p^d u$ by Lemma 2.4(a). And for any such g, $|A_g|$ is given by Lemma 2.5(b). Now, for a fixed d, there are

$$\tfrac{1}{2}(p-1)p^{k-d-1}$$

choices for u that give distinct values for $p^d u$. Putting these results together, and replacing d by $2d$, we have

$$\begin{aligned} B_3 &= \sum_{d=0}^{[(k-1)/2]} \frac{1}{2}(p-1)p^{2k-d-1}\left(2p^{k-2d}\left(\frac{p^{2d+1}+1}{p+1}\right) - 1\right) \\ &= \frac{p^{2k-1}(p-1)}{2(p+1)} \sum_{d=0}^{[(k-1)/2]} p^{-d}\bigl(2p^{k-2d}(p^{2d+1}+1) - p - 1\bigr). \end{aligned} \tag{2}$$

Summing the geometric sequences, we have

$$\sum_{d=0}^{[(k-1)/2]} p^{-d}(-p-1) = -(p+1)p^{-[(k-1)/2]}\Big(\frac{p^{[(k-1)/2]+1}-1}{p-1}\Big),$$

$$\sum_{d=0}^{[(k-1)/2]} p^{-d}\big(2p^{k-2d}(p^{2d+1}+1)\big) = 2p^{k-[(k-1)/2]+1}\Big(\frac{p^{[(k-1)/2]+1}-1}{p-1}\Big) +2p^{k-3[(k-1)/2]}\Big(\frac{p^{3[(k-1)/2]+3}-1}{p^3-1}\Big).$$

Substituting these equations in (2) and simplifying with the help of a computer algebra system, we obtain the desired expression. □

Proof of Theorem 1.2. There are q^m monic polynomials in $\mathbb{Z}_q[x]$ with degree m. Hence there are q^{m+2} pairs of monic polynomials (f, g) with $\deg f = m$, $\deg g = 2$. By Lemma 3.1, the probability that a pair of these polynomials is relatively prime is

$$1-\frac{|S_{m,2,q}|}{q^{m+2}} = 1-\frac{|T_{2,q}|}{q^4}.$$

Now $|T_{2,q}| = B_1+B_2+B_3$, with the values of B_i given by Lemma 2.5. Manipulating this expression with the help of a computer algebra system, one obtains

$$|T_{2,q}| = \frac{p^k}{2(p+1)}D,$$

where D equals the expression

$$2p^{2k+1}+2p^{2+k/2}(p-1)\Big(\frac{p^{3k/2}-1}{p^3-1}\Big)+p^{1+k/2}+3p^{k/2}+p^k-p-2$$

when k is even, and D equals

$$2p^{2k+1}+2(p-1)\Big(\frac{p^{2(k+1)}-p^{(k+1)/2}}{p^3-1}\Big)+3p^{(k+1)/2}+p^{(k-1)/2}+p^k-2p-1,$$

when k is odd. When k is even, algebraic manipulation shows

$$2p^{2k+1} = 2(p+1)p^{2k}-2p^{2k},$$

$$2p^{2+k/2}(p-1)\Big(\frac{p^{3k/2}-1}{p^3-1}\Big) = 2p^{2k}-2p^{2+k/2}+2(1-p^2)\sum_{i=1}^{k/2-1} p^{2k-3i},$$

$$p^{1+k/2}+3p^{k/2} = (p+1)(-2p^{1+k/2}+3p^{k/2})+2p^{2+k/2},$$

$$p^k-p-2 = -(p+1)\sum_{i=1}^{k-1}(-p)^i-2(p+1).$$

Adding both sides, the left hand side sums to D. With $f_k(x)$ defined as in the introduction, we then have

$$\frac{1}{2(p+1)}D = f_k(p).$$

Theorem 1.2 follows immediately for k even. Similar calculations establish it for k odd. □

Acknowledgment

The authors thank the anonymous referee for many helpful suggestions.

References

[Benjamin and Bennett 2007] A. T. Benjamin and C. D. Bennett, "The probability of relatively prime polynomials", *Math. Mag.* **80**:3 (2007), 196–202. MR 2008b:11036

[Corteel et al. 1998] S. Corteel, C. D. Savage, H. S. Wilf, and D. Zeilberger, "A pentagonal number sieve", *J. Combin. Theory Ser. A* **82**:2 (1998), 186–192. MR 99d:11111 Zbl 0910.05008

[Gao and Panario 2006] Z. Gao and D. Panario, "Degree distribution of the greatest common divisor of polynomials over $\mathbb{F}_q$", *Random Structures Algorithms* **29**:1 (2006), 26–37. MR 2008k:60020 Zbl 1099.11072

[Gouvêa 1997] F. Q. Gouvêa, *p-adic numbers*, 2nd ed., Universitext, Springer, Berlin, 1997. MR 98h:11155 Zbl 0874.11002

[Hou and Mullen 2009] X.-D. Hou and G. L. Mullen, "Number of irreducible polynomials and pairs of relatively prime polynomials in several variables over finite fields", *Finite Fields Appl.* **15**:3 (2009), 304–331. MR 2010c:11146 Zbl 05554713

[Reifegerste 2000] A. Reifegerste, "On an involution concerning pairs of polynomials over $\mathbf{F}_2$", *J. Combin. Theory Ser. A* **90**:1 (2000), 216–220. MR 2001a:11196 Zbl 1010.11068

Received: 2009-12-11 Revised: 2010-04-25 Accepted: 2010-04-26

hagedorn@tcnj.edu
The College of New Jersey, Department of Mathematics and Statistics, P.O. Box 7718, Ewing, NJ 08628, United States

hatley@math.umass.edu
The College of New Jersey, Department of Mathematics and Statistics, P.O. Box 7718, Ewing, NJ 08628, United States
Department of Mathematics and Statistics,
University of Massachusetts at Amherst, Amherst, MA 01003

𝔾-planar abelian groups

Andrea DeWitt, Jillian Hamilton, Alys Rodriguez and Jennifer Daniel

(Communicated by Scott Chapman)

For a group G with generating set $S = \{s_1, s_2, \dots, s_k\}$, the 𝔾-graph of G, denoted by $\Gamma(G, S)$, is the graph whose vertices are distinct cosets of $\langle s_i \rangle$ in G. Two distinct vertices are joined by an edge when the set intersection of the cosets is nonempty. In this paper, we explore the planarity of $\Gamma(G, S)$.

1. Introduction

Let G be a group with a generating set $S = \{s_1, \dots, s_k\}$. We say that the subset $T_{\langle s_i \rangle} \subset G$ is a *left transversal* for the subgroup $\langle s_i \rangle$ of G if $\{x\langle s_i \rangle \mid x \in T_{\langle s_i \rangle}\}$ is precisely the set of all left cosets of $\langle s_i \rangle$ in G. As in [Bauer et al. 2008], we associate with (G, S) a simple graph $\Gamma(G, S)$ with vertex set $V(\Gamma(G, S)) = \{x_j\langle s_i \rangle \mid x_j \in T_{\langle s_i \rangle}\}$. Two distinct vertices $x_j\langle s_i \rangle$ and $x_l\langle s_k \rangle$ in $V(\Gamma(G, S))$ are joined by an edge if $x_j\langle s_i \rangle \cap x_l\langle s_k \rangle$ is nonempty. The edge set, $E(\Gamma(G, S))$, consists of pairs $(x_j\langle s_i \rangle, x_l\langle s_k \rangle)$. $\Gamma(G, S)$ defined this way has no multiedge or loop.

Let $V_i = \{x_j\langle s_i \rangle \mid x_j \in T_{s_i}\}$. Then $V = \bigcup_{i=1}^{k} V_i$. The number of vertices in V_i is simply the order of G divided by the order of s_i which is the index of $\langle s_i \rangle$ in G, denoted $[G : \langle s_i \rangle]$. The minimum number of elements required to generate a finite group G is called the *rank of* G. A *minimal generating set for* G is a subset $S = \{s_1, \dots, s_k\}$ such that $G = \langle S \rangle$, where k is the rank of G. This concept is not to be confused with nonredundancy. A *nonredundant* set of generators is a set S such that S generates all of G, that is, $\langle S \rangle = G$, but no proper subset of S generates all of G.

The main object of this paper is to explore the planarity of $\Gamma(G, S)$.

Definition 1.1. A group G is 𝔾-planar if there exists a generating set S such that the graph, $\Gamma(G, S)$, is a planar graph.

We recall a fundamental criterion for the 𝔾-planarity of a group:

MSC2000: 05C25, 20F05.

Keywords: groups, graphs, generators.

A. Rodriguez and J. Daniel are partially supported by the MAA under its National Research Experience for Undergraduates Program, which is funded by the National Science Foundation, the National Security Agency, and the Moody Foundation.

Theorem 1.2 (Wagner). *A finite graph is planar if and only if it does not have K_5 or $K_{3,3}$ as a minor.*

2. Examples of $\mathbb{G}$-planar groups

The next two theorems give us two classes of $\mathbb{G}$-planar groups.

Theorem 2.1. *All cyclic groups are $\mathbb{G}$-planar.*

Proof. Let G be a cyclic group. Since G is cyclic, there exists an element $b \in G$ such that $\langle b \rangle = G$. Let $S = \{b\}$ be the generating set of G. Then $\Gamma(G, S)$ contains only one vertex and $\Gamma(G, S)$ is a planar graph. Therefore G is a $\mathbb{G}$-planar group. □

For the dihedral group, D_n, let r be a rotation of $360°/n$ and let f be any reflection.

Proposition 2.2. *For $S = \{f, rf\}$, the graph $\Gamma(G, S)$ of the dihedral group D_n is the cycle of length $2n$, C_{2n}.*

Proof. Write

$$V_1 = \{\langle f\rangle, r\langle f\rangle, r^2\langle f\rangle, \dots, r^{n-1}\langle f\rangle\},$$
$$V_2 = \{\langle rf\rangle, r\langle rf\rangle, r^2\langle rf\rangle, \dots, r^{n-1}\langle rf\rangle\}.$$

Since f and rf are both reflections, their composition is a rotation. Denote this rotation by r^m.

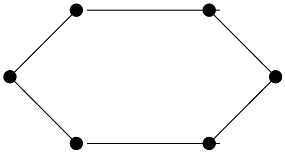
Choose a vertex from V_1, $r^s\langle f\rangle$. Since

$$r^s \in r^s\langle f\rangle \cap r^s\langle rf\rangle,$$

the edge $(r^s\langle f\rangle, r^s\langle rf\rangle)$ is in E. Now we need to show that there is another edge between $r^s\langle f\rangle$ and V_2. By simple calculation, we have $r^s f = r^{(s+m) \bmod n} rf$; moreover $(r^s\langle f\rangle, r^{(s+m) \bmod n}\langle rf\rangle)$ is in E.

Therefore the degree of each vertex in V_1 is 2. By similar arguments, the degree of each vertex in V_2 is 2 and $\Gamma(G, S)$ is a cycle. □

Example 2.3. Let $G = D_3$ and $S = \{f, rf\}$. Then the $\mathbb{G}$-graph is the cycle C_6:



Theorem 2.4. *All dihedral groups are $\mathbb{G}$-planar.*

Proof. Let $G = D_n$ and $S = \{f, rf\}$. Since $\Gamma(G, S)$ is a cycle, $\Gamma(G, S)$ is a planar graph and G is a $\mathbb{G}$-planar group. □

From [DeWitt et al. ≥ 2010], we have a few other examples of $\mathbb{G}$-planar groups.

Example 2.5. The modular group M has presentation

$$\langle s, t \mid s^8 = t^2 = e, st = ts^5 \rangle.$$

Let $S = \{s, ts\}$. From [DeWitt et al. ≥ 2010], $\Gamma(M, S)$ is $K_{2,2}$. Therefore $\Gamma(M, S)$ is a planar graph and M is a $\mathbb{G}$-planar group.

Example 2.6. The quasihedral group QS has presentation

$$\langle s, t \mid s^8 = t^2 = e, st = ts^3 \rangle.$$

Let $S = \{s, ts\}$. From [DeWitt et al. ≥ 2010], $\Gamma(QS, S)$ is $K_{2,4}$. Therefore $\Gamma(QS, S)$ is a planar graph and QS is a $\mathbb{G}$-planar group.

Recall that the generalized quaternion group Q_{2^n} has presentation

$$\langle s, t \mid s^{2^{n-1}} = e, s^{2^{n-2}} = t^2, tst^{-1} = s^{-1} \rangle.$$

Theorem 2.7. *The generalized quaternion group Q_{2^n} is $\mathbb{G}$-planar.*

Proof. Let $G = Q_{2^n}$ and $S = \{ts^k, ts^m\}$, where k is odd and m is even. $\Gamma(G, S)$ is a bipartite connected graph with every vertex of degree 2 [DeWitt et al. ≥ 2010]. Therefore, $\Gamma(G, S)$ is a cycle and Q_{2^n} is $\mathbb{G}$-planar. □

3. Finite abelian groups

The fundamental theorem of finite abelian groups tells us that every finite abelian group of rank k is isomorphic to a direct product of cyclic groups of prime-power order, that is, $G \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$. A *standard generating set for G* is a subset $S = \{s_1, \ldots, s_k\}$ such that $G = \langle s_1 \rangle \times \cdots \times \langle s_k \rangle$. Let G be an abelian group with standard generating set $S = \{s_1, \ldots, s_k\}$, then G is isomorphic to

$$\mathbb{Z}_{|s_1|} \times \mathbb{Z}_{|s_2|} \times \cdots \times \mathbb{Z}_{|s_k|}.$$

From Theorem 2.1, we know that all finite abelian groups with 1 generator are $\mathbb{G}$-planar. We now consider three cases: finite abelian groups with 4 or more generators, 3 generators or 2 generators.

Let G be a group with generating set S. There exists a subset of S, S', that is nonredundant and generates G. From [Bretto and Gillibert 2004], $\Gamma(G, S')$ is necessarily a subgraph of $\Gamma(G, S)$. If $\Gamma(G, S')$ is not a planar graph, then $\Gamma(G, S)$ is not planar. Therefore, it is only necessary to consider generating sets that are nonredundant.

Example 3.1. Let $G = \mathbb{Z}_2 \times \mathbb{Z}_6$ and $S = \{(1, 0), (0, 0), (0, 2), (0, 3), (0, 4)\}$. The subset $S' = \{(1, 0), (0, 2), (0, 3)\}$ of S is a nonredundant generating set of G. The set $S'' = \{(1, 0), (0, 1)\}$ is a minimal generating set of G that is also nonredundant.

Lemma 3.2. *Let G be a finite abelian group and let $S = \{s_1, s_2, s_3, \dots, s_k\}$ be a nonredundant generating set, then $|s_i| \geq 2$ for all i.*

Proof. Assume $|s_i| < 2$. Then $|s_i| = 1$ and $\langle s_i \rangle = \{e\}$. Therefore s_i is not needed to generate G and $S \setminus \{s_i\}$ generates G. This is a contradiction. Therefore, $|s_i| \geq 2$. □

Finite abelian groups G with* 4 *or more generators.

Lemma 3.3. *Let G be a finite abelian group and let $S = \{s_1, s_2, s_3, s_4, \dots, s_k\}$ be a nonredundant generating set of G with $k \geq 4$. Consider the subgroup H of G that is generated by $S' = \{s_1, s_2, s_3, s_4\}$. The vertices $\langle s_1 \rangle$, $\langle s_2 \rangle$, $\langle s_3 \rangle$, $\langle s_4 \rangle$, $s_1\langle s_2 \rangle$, $s_2\langle s_1 \rangle$, $s_2\langle s_3 \rangle$, $s_3\langle s_2 \rangle$, $s_3\langle s_4 \rangle$, $s_4\langle s_3 \rangle$ of $\Gamma(H, S')$ are all unique.*

Proof. To see that each of these vertices is unique, assume $\langle s_1 \rangle, s_2\langle s_1 \rangle \in V_1$ are not distinct, that is, $\langle s_1 \rangle = s_2\langle s_1 \rangle$. So there exists $k \in \mathbb{Z}^+$ such that $s_2 = s_1^k$ which contradicts the fact that S is a nonredundant generating set of G. The proofs of the other cases are similar. □

Theorem 3.4. *Let G be a finite abelian group and let $S = \{s_1, s_2, s_3, s_4, \dots, s_k\}$ be a nonredundant generating set of G with $k \geq 4$. Then $\Gamma(G, S)$ is not a planar graph.*

Proof. Consider the subgroup H of G generated by $S' = \{s_1, s_2, s_3, s_4\}$. Define a contraction Γ of $\Gamma(H, S')$ in this way: Let $\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_4, \overline{V}_5 \in V(\Gamma)$ with

$$\{\langle s_1 \rangle\} = \overline{V}_1, \quad \{\langle s_2 \rangle\} = \overline{V}_2, \quad \{\langle s_3 \rangle\} = \overline{V}_3, \quad \{\langle s_4 \rangle\} = \overline{V}_4,$$
$$\{s_1\langle s_2 \rangle, s_2\langle s_1 \rangle, s_2\langle s_3 \rangle, s_3\langle s_2 \rangle, s_3\langle s_4 \rangle, s_4\langle s_3 \rangle\} = \overline{V}_5.$$

Then, $e \in (\overline{V}_1 \cap \overline{V}_2)$, $e \in (\overline{V}_1 \cap \overline{V}_3)$, $e \in (\overline{V}_1 \cap \overline{V}_4)$, $s_1 \in (\overline{V}_1 \cap \overline{V}_5)$, $e \in (\overline{V}_2 \cap \overline{V}_3)$, $e \in (\overline{V}_2 \cap \overline{V}_4)$, $s_2 \in (\overline{V}_2 \cap \overline{V}_5)$, $e \in (\overline{V}_3 \cap \overline{V}_4)$, $s_3 \in (\overline{V}_3 \cap \overline{V}_5)$, and $s_4 \in (\overline{V}_4 \cap \overline{V}_5)$. Then $(\overline{V}_i, \overline{V}_j) \in E(\Gamma)$ for all $i \neq j$ and $\Gamma = K_5$. So, $\Gamma(H, S')$ has K_5 as a minor and $\Gamma(H, S')$ is not planar. From [Bretto et al. 2005], $\Gamma(H, S')$ is a subgraph of $\Gamma(G, S)$. Therefore, $\Gamma(G, S)$ is not a planar graph. □

Corollary 3.5. *Let G be a finite abelian group of rank* 4 *or more. Then G is not $\mathbb{G}$-planar.*

Finite abelian groups G with* 3 *generators.

Example 3.6. Let $G = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ with standard generating set

$$S = \{s_1, s_2, s_3\} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}.$$

The graph $\Gamma(G, S)$, illustrated in Figure 1, is a planar graph; hence G is a $\mathbb{G}$-planar group.

Next we show that this example is the only abelian group of rank three that is $\mathbb{G}$-planar.

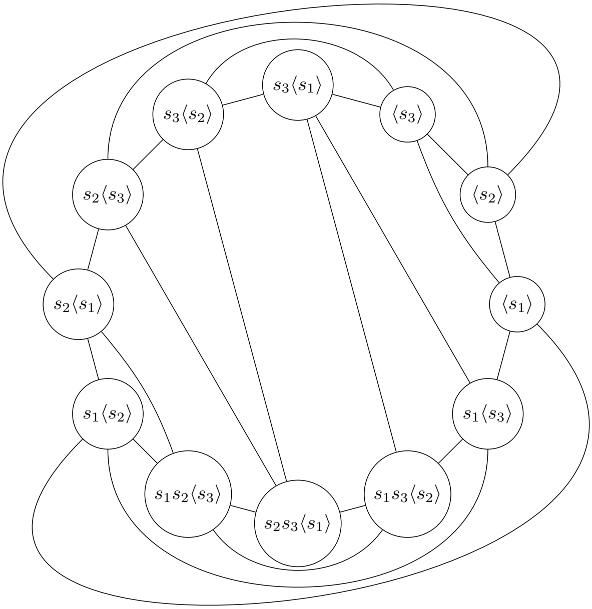


Figure 1. The graph $\Gamma(G, S)$, with $G = {\mathbb{Z}_2}^3$ and $S = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

Lemma 3.7. *Let G be a finite abelian group with nonredundant generating set $S = \{s_1, s_2, s_3\}$ such that $|s_i| \geq 3$ for at least one i. Then the graph $\Gamma(G, S)$ contains at least 16 vertices.*

Proof. Without loss of generality, assume that $|s_3| \geq 3$. There are at least 6 vertices in V_1. They are $\langle s_1\rangle$, $s_2\langle s_1\rangle$, $s_3\langle s_1\rangle$, $s_2s_3\langle s_1\rangle$, $s_3^2\langle s_1\rangle$, $s_2s_3^2\langle s_1\rangle$. To see that each of these vertices is unique, assume $\langle s_1\rangle, s_2s_3\langle s_1\rangle \in V_1$ are not distinct, that is, $\langle s_1\rangle = s_2s_3\langle s_1\rangle$. So there exists $k \in \mathbb{Z}^+$ such that $s_2s_3 = s_1^k$ which contradicts the fact that S is a nonredundant generating set of G. The proofs of the other cases are similar.

Likewise, there are at least 6 unique vertices in V_2 and 4 unique vertices on V_3. They are $\langle s_2\rangle$, $s_1\langle s_2\rangle$, $s_3\langle s_2\rangle$, $s_1s_3\langle s_2\rangle$, $s_3^2\langle s_2\rangle$, $s_1s_3^2\langle s_2\rangle$ and $\langle s_3\rangle$, $s_1\langle s_3\rangle$, $s_2\langle s_3\rangle$, $s_1s_2\langle s_3\rangle$. □

Theorem 3.8. *Let G be a finite abelian group with nonredundant generating set $S = \{s_1, s_2, s_3\}$ such that $|s_i| \geq 3$ for at least one i. Then $\Gamma(G, S)$ is not a planar graph.*

Proof. Define a contraction Γ of $\Gamma(G, S)$ by setting

$$\begin{aligned} \overline{V}_1 &= \{\langle s_1\rangle, \langle s_2\rangle\}, & \overline{V}_2 &= \{s_1\langle s_2\rangle, s_1s_2\langle s_3\rangle, s_1s_3\langle s_2\rangle\}, \\ \overline{V}_3 &= \{s_1\langle s_3\rangle, s_3^2\langle s_1\rangle, s_3^2\langle s_2\rangle\}, & \overline{V}_4 &= \{\langle s_3\rangle, s_3\langle s_2\rangle, s_3\langle s_1\rangle, s_2s_3\langle s_1\rangle\}, \\ \overline{V}_5 &= \{s_2\langle s_1\rangle, s_2\langle s_3\rangle, s_2s_3^2\langle s_1\rangle, s_1s_3^2\langle s_2\rangle\}. \end{aligned}$$

Then

$$\begin{aligned} s_1 &\in (\overline{V}_1 \cap \overline{V}_2), & s_1 &\in (\overline{V}_1 \cap \overline{V}_3), \\ e &\in (\overline{V}_1 \cap \overline{V}_4), & s_2 &\in (\overline{V}_1 \cap \overline{V}_5), \\ s_1 &\in (\overline{V}_2 \cap \overline{V}_3), & s_1 s_2 s_3 &\in (\overline{V}_2 \cap \overline{V}_4), \\ s_1 s_2 &\in (\overline{V}_2 \cap \overline{V}_5), & s_3^2 &\in (\overline{V}_3 \cap \overline{V}_4), \\ s_3^2 s_2 &\in (\overline{V}_3 \cap \overline{V}_5), & s_2 s_3 &\in (\overline{V}_4 \cap \overline{V}_5). \end{aligned}$$

It follows that $(\overline{V}_i, \overline{V}_j) \in E(\Gamma)$ for all $i \neq j$ and $\Gamma = K_5$. So, $\Gamma(G, S)$ has K_5 as a minor and is not a planar graph. □

Corollary 3.9. *Let G be a finite abelian group of rank 3 such that $G \not\cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. Then G is not a $\mathbb{G}$-planar group.*

Finite abelian groups G with 2 generators. Since we have results for groups of rank 1 and for groups of rank 3 or more, the only case left to consider is that of groups of rank 2. Notice that any finite abelian group of rank 2 is isomorphic to the direct product $\mathbb{Z}_m \times \mathbb{Z}_n$ with $gcd(m, n) \neq 1$.

Lemma 3.10. *Let G be a finite abelian group of rank 2 and let S be a nonredundant generating set of G. If $|S| \geq 3$, then $\Gamma(G, S)$ is not a planar graph.*

Proof. If $|S| > 3$, then $\Gamma(G, S)$ is not planar by Theorem 3.4. Assume that $|S| = 3$, that is, $S = \{s_1, s_2, s_3\}$ and that $|s_i| < 3$ for $i = 1, 2, 3$. Since S is nonredundant $|s_i| > 1$ and therefore $|s_i| = 2$ for $i = 1, 2, 3$. Consider the subset

$$H = \langle s_1 \rangle \langle s_2 \rangle = \{hk \mid h \in \langle s_1 \rangle, k \in \langle s_2 \rangle\} = \{e, s_1, s_2, s_1 s_2\}$$

of G. Since G is abelian, this subset is a subgroup. Now consider the subset

$$K = H \langle s_3 \rangle = \{hk \mid h \in H, k \in \langle s_3 \rangle\} = \{e, s_1, s_2, s_1 s_2, s_3, s_1 s_3, s_2 s_3, s_1 s_2 s_3\}$$

of G. Again K is necessarily a subgroup of G.

Now assume that $g \in G$. Since S generates G, there exists n, m, l such that $g = s_1^n s_2^m s_3^l$. Since the order of each generator is 2, n, m, l are congruent to 0 or 1 modulo 2 and $g \in K$. Therefore $G = K$. Since the order of each element in G is two, $G \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. This is a contradiction since G is a group of rank 2. Therefore, $|s_i| \geq 3$ for at least one i and by Theorem 3.8 the graph, $\Gamma(G, S)$, is not planar. □

Theorem 3.11. *Let G be a finite abelian group of rank 2. G is $\mathbb{G}$-planar if and only if $G \cong \mathbb{Z}_2 \times \mathbb{Z}_k$, for some $k \in \mathbb{N}$.*

Proof. ($\Leftarrow$) Let $G \cong \mathbb{Z}_2 \times \mathbb{Z}_k$ and let $\Gamma(\mathbb{Z}_2 \times \mathbb{Z}_k, S)$ be the associated $\mathbb{G}$-graph of $\mathbb{Z}_2 \times \mathbb{Z}_k$ with $S = \{(1, 0), (0, 1)\}$. There exist an isomorphism $\phi : \mathbb{Z}_2 \times \mathbb{Z}_k \to G$. Let $(x, y) \in \mathbb{Z}_2 \times \mathbb{Z}_k$. There exists a, b such that $(x, y) = a(1, 0) + b(0, 1)$. Then $\phi(x, y) = \phi(a(1, 0) + b(0, 1)) = a\phi(1, 0) \oplus b\phi(0, 1)$. So $\phi(S) = \{\phi(1, 0), \phi(0, 1)\}$

Rank	Group	Planarity
1	all G	planar
2	$G \cong \mathbb{Z}_2 \times \mathbb{Z}_k$ $G \ncong \mathbb{Z}_2 \times \mathbb{Z}_k$	planar not planar
3	$G \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ $G \ncong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$	planar not planar
4 or more	all G	not planar

Table 1. 𝔾-planarity of finite abelian groups.

generates G. $\Gamma(\mathbb{Z}_2 \times \mathbb{Z}_k, \mathrm{S})$ is $K_{k,2}$, so $K_{k,2} \cong \Gamma(G, \phi(S))$. Since $K_{k,2}$ is planar, $\Gamma(G, \phi(S))$ is planar. Therefore G is 𝔾-planar.

($\Rightarrow$) Let G be a finite abelian 𝔾-planar group of rank 2 and let S be a generating set such that $\Gamma(G, S)$ is a planar graph. From Lemma 3.10, $|S| = 2$, that is, $S = \{s_1, s_2\}$.

Case 1. Assume that $|s_1| = 2$. Let $|G| = n$, $|V_1| = [G : \langle s_1\rangle] = n/2$. So

$$V_1 = \{\langle s_1\rangle, s_2\langle s_1\rangle, s_2^2\langle s_1\rangle, \cdots, s_2^{n/2-1}\langle s_1\rangle\},$$

and the elements of G are of the form

$$s_2, s_2^2, \ldots, s_2^{n/2-1}, e \quad \text{and} \quad s_1s_2, s_1s_2^2, \ldots, s_1s_2^{n/2-1}, s_1.$$

Therefore $|s_2| = n/2$ and G is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_{n/2}$.

Case 2. Assume that $|s_1|, |s_2| > 2$. Consider the vertex induced subgraph generated by the six vertices $\langle s_1\rangle, s_2\langle s_1\rangle, s_2^2\langle s_1\rangle, \langle s_2\rangle, s_1\langle s_2\rangle, s_1^2\langle s_2\rangle$. This graph is $K_{3,3}$. Since this subgraph is not planar, $\Gamma(G, S)$ is not planar. This contradicts the supposition that S is a generating set such that $\Gamma(G, S)$ is a planar graph. Therefore, if G is 𝔾-planar, then $G \cong \mathbb{Z}_2 \times \mathbb{Z}_k$. □

Table 1 summarizes the results for all finite abelian groups.

References

[Bauer et al. 2008] C. M. Bauer, C. K. Johnson, A. M. Rodriguez, B. D. Temple, and J. R. Daniel, "Paths and circuits in 𝔾-graphs", *Involve* **1**:2 (2008), 135–144. MR 2009j:05108 Zbl 1146.05306

[Bretto and Gillibert 2004] A. Bretto and L. Gillibert, "Graphical and computational representation of groups", pp. 343–350 in *Computational science* (ICCS 2004) (Kraków, 2004), edited by M. Bubak et al., Lecture Notes in Comput. Sci. **3039**, Springer, Berlin, 2004. MR 2233213 Zbl 1102.68735

[Bretto et al. 2005] A. Bretto, L. Gillibert, and B. Laget, "Symmetric and semisymmetric graphs construction using G-graphs", pp. 61–67 in *Algorithms and Computation: 16th International Symposium, ISAAC* (Sanya, Hainan, China), edited by X. Deng and D.-Z. Du, ACM, New York, 2005. MR 2280530

[DeWitt et al. $\geq$ 2010] A. DeWitt, A. Rodriguez, and J. Daniel, "Paths and circuits in $\mathbb{G}$-graphs of certain non-abelian groups", *Furman Univ. Electr. J. Undergrad. Math.*. To appear.

Received: 2010-04-26 Revised: 2010-06-17 Accepted: 2010-06-24

aldewitt@my.lamar.edu *Lamar University, Department of Mathematics,
Beaumont, TX 77710, United States*

jkhamilton1@my.lamar.edu *Lamar University, Department of Mathematics,
Beaumont, TX 77710, United States*

amrodriguez1@my.lamar.edu *Lamar University, Department of Mathematics,
Beaumont, TX 77710, United States*

Jennifer.Daniel@lamar.edu *Lamar University, Department of Mathematics,
Beaumont, TX 77710, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the Involve website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *Involve* are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

involve

2010 vol. 3 no. 2

Recursive sequences and polynomial congruences 129
J. Larry Lehman and Christopher Triola

The Gram determinant for plane curves 149
Józef H. Przytycki and Xiaoqi Zhu

The cardinality of the value sets modulo n of x^2+x^{-2} and x^2+y^2 171
Sara Hanrahan and Mizan Khan

Minimal k-rankings for prism graphs 183
Juan Ortiz, Andrew Zemke, Hala King, Darren Narayan and Mirko Horňák

An unresolved analogue of the Littlewood Conjecture 191
Clarice Ferolito

Mapping the discrete logarithm 197
Daniel Cloutier and Joshua Holden

Linear dependency for the difference in exponential regression 215
Indika Sathish and Diawara Norou

The probability of relatively prime polynomials in $\mathbb{Z}_{p^k}[x]$ 223
Thomas R. Hagedorn and Jeffrey Hatley

$\mathbb{G}$-planar abelian groups 233
Andrea DeWitt, Jillian Hamilton, Alys Rodriguez and Jennifer Daniel