involve

a journal of mathematics

mathematical sciences publishers

2010 vol. 3, no. 3

involve

pjm.math.berkeley.edu/involve

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

INVOLVE 3:3(2010)

Gracefulness of families of spiders

Patrick Bahls, Sara Lake and Andrew Wertheim

(Communicated by Jerrold Griggs)

We say that a tree is a *spider* if it has at most one branch point. We prove the existence of a family of graceful labelings for spiders all of whose legs are equal in length.

1. Introduction

Let $G = (V, E)$ be a (simple, undirected) graph. A *labeling* of G is a map from the set V of vertices to the set of nonnegative integers. A labeling ϕ induces a labeling on the edge set E by assigning to $e = \{u, v\}$ the value $\phi(e) = |\phi(u) - \phi(v)|$.

A labeling is said to be *graceful* if its labels take values in $\{0, 1, \ldots, |V| - 1\}$, it has no repeated labels, and its induced edge labeling has no repeated labels.

A graph is graceful if there is some graceful labeling of its vertices. Graceful labelings were first defined by Rosa as he considered problems involving decompositions of graphs; see [Rosa 1967], in which various sorts of labelings are defined. Golomb [1972] was the first to use the term *graceful labeling*.

There is a long-standing conjecture that every tree — that is, every connected acyclic graph — is graceful. Known as the Ringel–Kotzig conjecture, it seems to have first been published as Problem 25, p. 162 in a collection of open problems in [Fiedler 1964]. See [Edwards and Howard 2006; Gallian 1997–2009] for more information on this conjecture and hundreds of related results. We note that proofs of gracefulness for general classes of trees are hard to come by.

We call the graph T a *spider* if it has at most one branch point — that is, at most one vertex v such that the degree $d(v)$ satisfies $d(v) \geq 3$. Let v^* denote the unique branch point of a spider T, if this point exists. We call this point the *center* of the graph T. A *leg* of the spider T is any one of the paths from v^* to a leaf of T. We will prove the following result in Section 2:

Theorem 1. *Let T be a spider with l legs, each of which has length in $\{m, m+1\}$ for some $m \geq 1$. Then T is graceful.*

MSC2000: 05C78.

Keywords: graceful labeling, graph labeling, tree.

Theorem 1 is not a new result. It follows from [Poljak and Sûra 1982], but our proof also shows gracefulness for any tree formed by appending an extra leg of any length to an odd-legged spider with legs of lengths in $\{m, m+1\}$. A generalization of the construction, given in Section 3, leads to further interesting labelings: specifically, for spiders having an odd number of legs, all of equal length m, we construct for each positive divisor d of m a graceful labeling associated with d. This construction can be used to generate graceful labelings of many trees that are not spiders, as shown in [Bahls 2008].

2. Proof of the main theorem

We may assume that $l \geq 3$, as otherwise T is a path, which is known to be graceful. (For example, see [Aldred et al. 2003], in which an estimate is obtained for the number of graceful labelings on a path of a given length.)

Proof of Theorem 1 for l odd. Let $l = l_0 + l_1$, where l_i is the number of legs of length $m+i$ for $i \in \{0, 1\}$. Note that T has $n+1 = lm+l_1+1$ vertices, to be labeled by the set $\{0, 1, \ldots, n\}$. Label the legs by $L_1, L_2, \ldots, L_l$ so that $L_1, \ldots, L_{l_1}$ have length $m+1$ and $L_{l_1+1}, \ldots, L_l$ have length m. Let v^* denote the branch point of T and denote by $v_{i,j}$ the vertex in L_i of distance j from v^*.

Let ϕ be the labeling defined as follows:

(i) $\phi(v^*) = 0$;

(ii) if i and j are both odd,

$$\phi(v_{i,j}) = n - \frac{i-1}{2} - \frac{(j-1)l}{2};$$

(iii) if i and j are both even;

$$\phi(v_{i,j}) = n - \frac{l-1}{2} - \frac{i}{2} - \frac{(j-2)l}{2};$$

(iv) if i is even and j is odd,

$$\phi(v_{i,j}) = \frac{i}{2} + \frac{(j-1)l}{2};$$

(v) if i is odd and j is even,

$$\phi(v_{i,j}) = \frac{l-1}{2} + \frac{i+1}{2} + \frac{(j-2)l}{2}.$$

The labeling ϕ places 0 at the spider's center and, traversing the longer legs first, alternates between the highest and the lowest remaining unused labels, spiraling away from the center. This is illustrated in Figure 1, in which $l_0 = 2$, $l_1 = 3$, and $m = 4$.

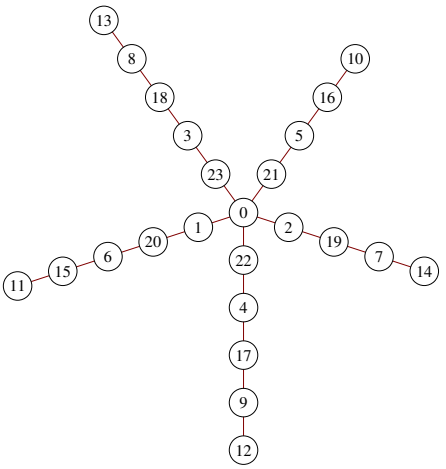


Figure 1. The labeling ϕ for $l_0 = 2$, $l_1 = 3$, and $m = 4$.

To help compute the induced edge labels, we note that the local maxima of ϕ occur at $v_{i,j}$ for which i and j have the same parity — that is, $i \equiv j \pmod 2$, For such i and j, we have

$$\phi(v_{i,j}) - \phi(v_{i,j+1}) = n - \frac{l-1}{2} - i + (1-j)l > 0, \tag{1}$$

$$\phi(v_{i,j}) - \phi(v_{i,j-1}) = n - \frac{l-1}{2} - i + (2-j)l > 0. \tag{2}$$

Suppose, to obtain a contradiction, that there are two distinct edges that share the same label. By considering the indexes of the vertices at both ends end of these edges, we see that we can choose distinct pairs of indexes (i, j) and (i', j') such that i and j have the same parity, i' and j' likewise have the same parity, and an edge incident on $v_{i,j}$ shares the same label as a different edge incident on $v_{i',j'}$, that is, one of these three cases occur:

$$\phi(v_{i,j}) - \phi(v_{i,j+1}) = \phi(v_{i',j'}) - \phi(v_{i',j'+1}), \tag{3}$$

$$\phi(v_{i,j}) - \phi(v_{i,j+1}) = \phi(v_{i',j'}) - \phi(v_{i',j'-1}), \tag{4}$$

$$\phi(v_{i,j}) - \phi(v_{i,j-1}) = \phi(v_{i',j'}) - \phi(v_{i',j'-1}). \tag{5}$$

Consider first the case where (3) holds. From (1), we obtain $i - i' + (j - j')l = 0$, which shows that $j \neq j'$, since otherwise $i = i'$ as well, contrary to the assumption that $(i, j) \neq (i', j')$. We therefore can write

$$l = \frac{i - i'}{j' - j}.$$

Thus $|i-i'|<l$ and $|j-j'|\geq 1$, and

$$l=\left|\frac{i-i'}{j'-j}\right|<\frac{l}{1}=l,$$

a contradiction.

Similar contradictions arise when (4) or (5) hold. Thus no two distinct edges bear the same labels, and ϕ is graceful. □

Proof of Theorem 1 for l even. Without loss of generality assume L_l is a leg of length m. Remove it, resulting in a tree T_0 with an odd number of legs, $l-1$. The construction above yields a graceful labeling ϕ_0 of T_0 such that $\phi_0(v^*)=0$. Let $|V(T_0)|=n'+1$. We define a new graceful labeling, ϕ_0', on T_0 by $\phi_0'(v)=n'-\phi_0(v)$ for all v.

Construct a new tree T_1 by appending a new vertex, w_1, to T_0's center. Define ϕ_1 on $V(T_1)$ by $\phi_1(w_1)=0$ and $\phi_1(v)=\phi_0'(v)+1$ for all $v\in V(T_0)$. Define ϕ_1' on T_1 by $\phi_1'(v)=n'+1-\phi_1(v)$ for all v; note that $\phi_1'(w_1)=n'+1$.

We now append a vertex w_2 to w_1 and construct graceful labelings ϕ_2 from ϕ_1', ϕ_2' from ϕ_2, and so forth, until we have reconstructed $L_l=\{w_1, w_2, \ldots, w_m\}$, recovering T. □

The argument in the case of l even actually shows this:

Theorem 2. *Let T be a spider with l legs, where l is even. Suppose each leg, except possibly one, has length in $\{m, m+1\}$ for some $m\geq 1$. Then T is graceful.*

3. A family of graceful labelings

Now assume that T is a spider with an odd number l of legs, each of length m. Let d be any fixed positive divisor of m; we define a graceful labeling ϕ_d corresponding to d.

We retain the notation $v_{i,j}$ from the previous section. Given a pair (i, j), set $t=\lceil j/d\rceil$ and $r=j-(t-1)d$. Roughly, t gives the "tier" of length d inside the i-th leg in which the vertex $v_{i,j}$ lies, and r gives its position relative to that tier. The value of $\phi_d(v_{i,j})$ will depend on the parity of each of d, i, t, and r, so we consider the vector $\vec{v}_{i,j}=(d, i, t, r)$ as an element of $\mathbb{Z}_2^4$ by reducing all coordinates modulo 2.

Let $\phi_d(v^*)=0$, as before. The following formula gives $\phi_d(v_{i,j})$:

(i) if $\vec{v}_{i,j}\in\{(0,1,1,1),(1,1,1,1)\}$,

$$\phi_d(v_{i,j})=ml-\frac{(t-1)ld}{2}-\frac{(i-1)d}{2}-\frac{r-1}{2};$$

(ii) if $\vec{v}_{i,j} \in \{(0,1,1,0),(1,1,1,0)\}$,

$$\phi_d(v_{i,j}) = \frac{(t-1)ld}{2} + \frac{(i-1)d}{2} + \frac{r}{2};$$

(iii) if $\vec{v}_{i,j} \in \{(0,0,1,1),(1,0,1,1)\}$,

$$\phi_d(v_{i,j}) = \frac{(t-1)ld}{2} + \frac{id}{2} - \frac{r-1}{2};$$

(iv) if $\vec{v}_{i,j} \in \{(0,0,1,0),(1,0,1,0)\}$,

$$\phi_d(v_{i,j}) = ml - \frac{(t-1)ld}{2} - \frac{id}{2} + \frac{r}{2};$$

(v) if $\vec{v}_{i,j} \in \{(1,1,0,1),(0,1,0,0)\}$,

$$\phi_d(v_{i,j}) = \left\lceil \frac{ld}{2} \right\rceil + \frac{(t-2)ld}{2} + \frac{(i-1)d}{2} + \left\lfloor \frac{r}{2} \right\rfloor;$$

(vi) if $\vec{v}_{i,j} \in \{(1,0,0,1),(0,0,0,0)\}$,

$$\phi_d(v_{i,j}) = ml - \left\lfloor \frac{ld}{2} \right\rfloor - \frac{(t-2)ld}{2} - \frac{id}{2} + \left\lfloor \frac{r}{2} \right\rfloor.$$

(vii) if $\vec{v}_{i,j} \in \{(1,1,0,0),(0,1,0,1)\}$,

$$\phi_d(v_{i,j}) = ml - \left\lceil \frac{ld}{2} \right\rceil - \frac{(t-2)ld}{2} - \frac{(i-1)d}{2} - \left\lceil \frac{r}{2} \right\rceil + 1;$$

(viii) if $\vec{v}_{i,j} \in \{(1,0,0,0),(0,0,0,1)\}$,

$$\phi_d(v_{i,j}) = \left\lfloor \frac{ld}{2} \right\rfloor + \frac{(t-2)ld}{2} + \frac{id}{2} - \left\lceil \frac{r}{2} \right\rceil + 1.$$

That this yields a graceful labeling can be proved in a manner similar to the proof of Theorem 1.

Like the labeling introduced in the proof of Theorem 1, this labeling proceeds by alternating between the greatest and least labels yet unused, spiraling outward from the center. Now, however, d vertices on each leg are labeled before proceeding to the next leg, and the direction in which the labeling proceeds within this length-d segment (inward or outward relative to the center) alternates from one leg to the next. An example is shown in Figure 2.

In the special case $d = 1$, we obtain the labeling constructed in the proof of Theorem 1. In this case $t = j$ and $r = 1$, so our labeling depends only on the parities of i and j, and indeed after reduction the corresponding formulas in the above list, namely (i), (iii), (v), and (vi), coincide precisely with those in the proof of Theorem 1.

The labelings ϕ_d have the property that the edges

$$\{v^*, v_{i,1}\}, \{v_{i,d}, v_{i,d+1}\}, \{v_{i,2d}, v_{i,2d+1}\}, \ldots, \{v_{i,m-d}, v_{i,m-d+1}\}$$

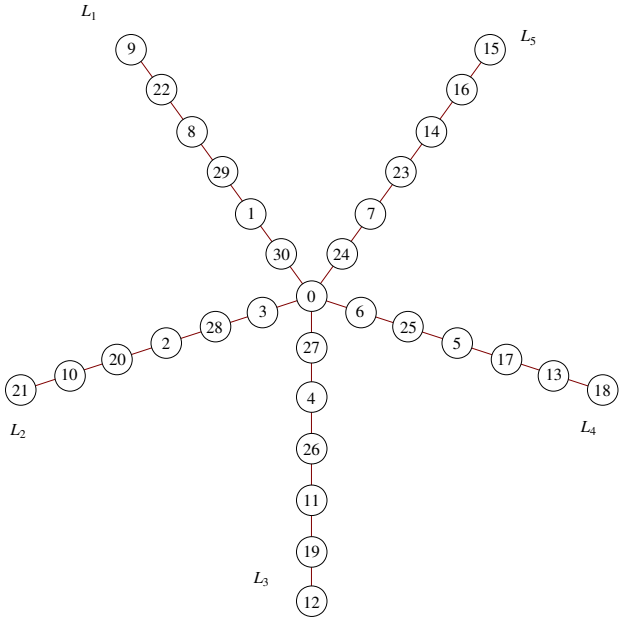


Figure 2. The labeling ϕ_d for $l = 5$, $m = 6$, and $d = 3$.

have labels divisible by d. This fact enables us to "deflate" the labeling ϕ_d and obtain a labeling ϕ'_d on the spider T' with l legs, each of length m/d. This new labeling is defined inductively as follows, spiraling outward from the center v' of T', where we denote by $v'_{i,j}$ the vertex in T' in position (i, j) as before and let $v_{i,0} = v^*$, $v'_{i,0} = v'$:

(i) $\phi'_d(v') = 0$;

(ii) $\phi'_d(v'_{i,1}) = \phi_d(v_{i,1})/d$;

(iii) assuming $\phi'_d(v'_{i,j})$ has been defined, let

$$\phi'_d(v'_{i,j+1}) = \phi'_d(v'_{i,j}) + (-1)^{l+j+1}\frac{\phi_d(\{v_{i,jd}, v_{i,jd+1}\})}{d}.$$

This process acts as an inverse to the process of edge subdivision considered in [Bahls 2008], in which each edge of a given gracefully labeled tree is subdivided a fixed number of times, yielding a new graph that can be gracefully labeled by making use of the labeling on the original tree.

References

[Aldred et al. 2003] R. E. L. Aldred, J. Širáň, and M. Širáň, "A note on the number of graceful labellings of paths", *Discrete Math.* **261**:1-3 (2003), 27–30. MR 2004a:05135 Zbl 1008.05132

[Bahls 2008] P. Bahls, "Generating graceful trees by subdivision", preprint, 2008, Available at http://facstaff.unca.edu/pbahls/papers/GracefulSubdivision.pdf.

[Edwards and Howard 2006] M. Edwards and L. Howard, "A survey of graceful trees", *Atlantic Electronic J. Math.* **1**:1 (2006), 5–30.

[Fiedler 1964] M. Fiedler (editor), *Theory of graphs and its applications* (Smolenice, 1963), Czechoslovak Acad. Sciences, Prague, and Academic Press, New York, 1964.

[Gallian 1997–2009] J. Gallian, "A dynamic survey of graph labeling", *Electron. J. Combin.* **DS6** (1997–2009). MR 99m:05141 Zbl 0953.05067

[Golomb 1972] S. W. Golomb, "How to number a graph", pp. 23–37 in *Graph theory and computing*, edited by R. C. Read and C. Berge, Academic Press, New York, 1972. MR 49 #4863 Zbl 0293.05150

[Poljak and Sûra 1982] S. Poljak and M. Sûra, "An algorithm for graceful labelling of a class of symmetrical trees", *Ars Combin.* **14** (1982), 57–66. MR 84d:05150 Zbl 0504.05029

[Rosa 1967] A. Rosa, "On certain valuations of the vertices of a graph", pp. 349–355 in *Internat. Sympos. Theory of Graphs* (Rome, 1966), Gordon and Breach, New York, 1967. MR 36 #6319 Zbl 0193.53204

Received: 2009-03-02 Revised: Accepted: 2010-07-25

pbahls@unca.edu *University of North Carolina, Asheville, Department of Mathematics, CPO #2350, One University Heights, Asheville, NC 28804-8511, United States*

salake@unca.edu *University of North Carolina, Asheville, Department of Mathematics, CPO #2350, One University Heights, Asheville, NC 28804-8511, United States*

ajwerthe@unca.edu *University of North Carolina, Asheville, Department of Mathematics, CPO #2350, One University Heights, Asheville, NC 28804-8511, United States*

INVOLVE 3:3(2010)

Rational residuacity of primes

Mark Budden, Alex Collins, Kristin Ellis Lea and Stephen Savioli

(Communicated by Filip Saidak)

The most natural extensions to the law of quadratic reciprocity are the rational reciprocity laws, described using the rational residue symbol. In this article, we provide a reciprocity law from which many of the known rational reciprocity laws may be recovered by picking appropriate primitive elements for subfields of $\mathbb{Q}(\zeta_p)$. As an example, a new generalization of Burde's law is provided.

1. Introduction

The law of quadratic reciprocity has played a central role in the development of number theory since Gauss published its first proof in 1801 (see [Lemmermeyer 2000] for the history of this important result). To state the law, assume that $a \in \mathbb{Z}$ is not divisible by an odd prime p and define the Legendre symbol by

$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{if } x^2 \equiv a \pmod{p} \text{ is solvable,} \\ -1 & \text{if not.} \end{cases}$$

Then if p and q are distinct odd primes, we have

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}.$$

The remainder of the 1800s and early 1900s saw many generalizations of this result to higher powers, culminating in class field theory, in which generalized reciprocity laws were established. Making such generalizations requires one to leave the realm of the integers, introducing rings of integers in algebraic number fields and primes within these rings. Hence, the study of reciprocity laws can serve as a great topic for students interested in learning about field extensions and Galois theory.

While class field theory has succeeded in capturing the true essence of the higher reciprocity laws, the extensions to the law of quadratic reciprocity that are the most accessible to students are the rational reciprocity laws. Such laws make use of the

MSC2000: primary 11A15; secondary 11R32, 11R18.

Keywords: reciprocity laws, ramification of prime ideals, cyclotomic fields.

Supported in part by an internal grant from Armstrong Atlantic State University and a CURM mini-grant funded through NSF grant DMS-0636648.

rational residue symbol, which only takes on the integer values ± 1 and is defined on rational primes. The simplicity of the rational residue symbol is much more tangible to students than the power residue symbol, making such laws an excellent starting point for students in algebraic number theory. Like the law of quadratic reciprocity, the statements are often elementary, but the proofs elucidate the utility of Galois theory and the ramification theory of prime ideals in algebraic number fields.

We begin with a description of the quadratic residue symbol and the rational residue symbol. Let K be an algebraic number field and N the norm map of K over $\mathbb{Q}$. Let $\mathfrak{p}$ be a prime ideal such that $\mathfrak{p} \nmid 2\mathcal{O}_K$, where $\mathcal{O}_K$ is the ring of integers in K. For every $\alpha \in \mathcal{O}_K - \mathfrak{p}$, define the quadratic residue symbol $\left(\frac{\alpha}{\mathfrak{p}}\right)$ by

$$\left(\frac{\alpha}{\mathfrak{p}}\right) \equiv \alpha^{(N(\mathfrak{p})-1)/2} \pmod{\mathfrak{p}}.$$

In the case where $K = \mathbb{Q}$, our definition agrees with the Legendre symbol on the generator of the prime ideal $\mathfrak{p} = p\mathbb{Z}$.

Now let $a \in \mathbb{Z}$ and p be an odd prime satisfying $(a, p) = 1$ such that

$$a^{(p-1)/n} \equiv 1 \pmod{p}.$$

Then the $2n$-th rational residue symbol $(a/p)_{2n}$ is defined by

$$\left(\frac{a}{p}\right)_{2n} \equiv a^{(p-1)/(2n)} \pmod{p}.$$

It is easily verified that this symbol only takes on the integer unit values ± 1. It should also be noted that it agrees with the $2n$-th power residue symbol $(a/\mathfrak{p})_{\mathbb{Q}(\zeta_{2n})}$, where $\mathfrak{p}$ is any prime ideal above p in $\mathbb{Q}(\zeta_{2n})$ and ζ_{2n} is the primitive $2n$-th root of unity $e^{\pi i/n}$.

An indispensable object used in the proofs of most reciprocity laws is the Galois group

$$\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}),$$

defined to be the group of all automorphisms $\mathbb{Q}(\zeta_p) \to \mathbb{Q}(\zeta_p)$ that fix $\mathbb{Q}$ pointwise (here, ζ_p is the primitive p-th root of unity $e^{2\pi i/p}$). By the fundamental theorem of Galois theory (see [Gallian 2010, Chapter 32], for instance), there is a one-to-one correspondence between the intermediate subfields of the extension $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ and the subgroups of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$. It is well known that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$$

is a cyclic group of order $p-1$. So, whenever $p \equiv 1 \pmod{m}$, there exists a unique subfield K_m of $\mathbb{Q}(\zeta_p)$ that satisfies $[K_m : \mathbb{Q}] = m$.

Lemmermeyer [1994] showed that when $p \equiv 1 \pmod 4$, specific choices of $A, B \in \mathbb{Z}$ so that $K_4 = \mathbb{Q}(\sqrt{A + B\sqrt{p}})$ result in the rational quartic reciprocity laws of Scholz [1934], Lehmer [1958; 1978], and Burde [1969]. His work simplified the all-encompassing rational quartic reciprocity law of Williams et al. [1985] as well as its simplification by Evans [1989]. The reader unfamiliar with these laws may consult Lehmer's survey article [Lehmer 1978] and [Lemmermeyer 2000] for the relevant background.

When extending the known rational quartic reciprocity laws, it is natural to look for analogues that involve the 2^t-th rational residue symbols $(p/q)_{2^t}$ and $(q/p)_{2^t}$ when $p \equiv q \equiv 1 \pmod{2^t}$ are distinct primes. Such a generalization of Burde's law was proved by Evans [1981], and Budden et al. [2007] recently proved such a generalization of Scholz's law. In Section 2, we follow the approach of [Budden et al. 2007] to prove a $2n$-th reciprocity law (Theorem 1), from which many of the known rational reciprocity laws can be recovered. The approach is similar to that of [Lemmermeyer 1994] in that it compares the factorization of the prime ideal $q\mathbb{Z}$ in $\mathbb{Q}(\zeta_p)$ to its factorization in K_{2n}. Additionally, the all-encompassing rational quartic law in this last reference may be viewed as a special case of the quartic version of the $2n$-th law presented here. Hence, all of the known rational quartic reciprocity laws may be recovered from Theorem 1.

Finally, as an application of Theorem 1, we give in Section 3 a 2^t-th generalization of Burde's law (Theorem 3), that differs from the known generalizations. In particular, our result is different from Williams' octic version of Burde's law [Williams 1976] when $t = 3$ (also proved independently by Wu [1975]), Leonard and Williams' sixteenth version of Burde's law when $t = 4$ [Leonard and Williams 1977], and Evans' 2^t-th generalization of Burde's law [Evans 1981]. Interesting results follow from comparing the variations.

2. A $2n$-th rational reciprocity law

Now assume that $p \equiv q \equiv 1 \pmod{2n}$ are distinct primes with $n \geq 1$ such that

$$\left(\frac{p}{q}\right)_n = \left(\frac{q}{p}\right)_n = 1.$$

Then the ideal $q\mathbb{O}_{K_n}$ factors into prime ideals as

$$q\mathbb{O}_{K_n} = \lambda_1 \lambda_2 \cdots \lambda_n,$$

with all of the λ_i distinct. We obtain the following reciprocity law.

Theorem 1. *Let $p \equiv q \equiv 1 \pmod{2n}$ be distinct primes with $n \geq 1$ and assume*

$$\left(\frac{p}{q}\right)_n = \left(\frac{q}{p}\right)_n = 1.$$

If $\beta \in \mathbb{O}_{K_n}$ is such that $K_{2n} = K_n(\sqrt{\beta})$, then $\left(\frac{q}{p}\right)_{2n} = \left(\frac{\beta}{\lambda}\right)$, where λ is any prime ideal above q in $\mathbb{O}_{K_n}$.

Proof. The cyclotomic polynomial $\Phi_p(x) = \prod_{k=1}^{p-1}(x - \zeta_p^k)$ splits over K_n, and we let $\varphi_p(x)$ be the irreducible factor

$$\varphi_p(x) = \prod_{\substack{1 \le r \le p-1 \\ (r/p)_n = 1}} (x - \zeta_p^r).$$

Since $\Phi_p(x) \in \mathbb{Z}[\zeta_p][x]$, it follows that $\varphi_p(x) \in \mathbb{O}_{K_n}$. Furthermore, it has degree $(p-1)/n$ and splits further over K_{2n} into $\varphi_p(x) = \psi_p(x) \cdot \tilde{\psi}_p(x)$, where

$$\psi_p(x) = \prod_{\substack{1 \le r \le p-1 \\ (r/p)_{2n} = 1}} (x - \zeta_p^r) \qquad \text{and} \qquad \tilde{\psi}_p(x) = \prod_{\substack{1 \le t \le p-1 \\ (t/p)_{2n} = -1 \\ (t/p)_n = 1}} (x - \zeta_p^t).$$

Define the polynomial $\vartheta(x) = \psi_p(x) - \tilde{\psi}_p(x) \in \mathbb{O}_{K_{2n}}[x]$ and consider the automorphism $\sigma_q \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$, defined by $\sigma_q(\zeta_p) = \zeta_p^q$. Since the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic, it has unique cyclic subgroups of orders dividing $p-1$, implying that

$$\mathrm{Gal}(\mathbb{Q}(\zeta_p)/K_n) \cong (\mathbb{Z}/p\mathbb{Z})^{\times n} \quad \text{and} \quad \mathrm{Gal}(\mathbb{Q}(\zeta_p)/K_{2n}) \cong (\mathbb{Z}/p\mathbb{Z})^{\times 2n}.$$

Under the assumption $(q/p)_n = 1$, the automorphism σ_q is contained in the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/K_n)$. Its restriction to K_{2n} must agree with either the identity automorphism $I \in \mathrm{Gal}(K_{2n}/K_n)$, or the nontrivial automorphism $\alpha(\sqrt{\beta}) = -\sqrt{\beta}$. It follows that

$$\sigma_q|_{K_{2n}} = I \iff (q/p)_{2n} = 1.$$

Since

$$\alpha(\sqrt{\beta}\ \vartheta(x)) = \sqrt{\beta}\ \vartheta(x)$$

and the coefficients in $\vartheta(x)$ come from $\mathbb{O}_{K_{2n}}$, every coefficient must be an element in $\mathbb{O}_{K_n}$ multiplied by $\sqrt{\beta}$ so that we can write

$$\vartheta(x) = \sqrt{\beta}\ \phi(x), \quad \text{for some } \phi(x) \in \mathbb{O}_{K_n}[x].$$

We have also assumed that $(p/q)_n = 1$, so that the ideal $q\mathbb{O}_{K_n}$ splits completely in $\mathbb{O}_{K_n}$ (i.e., $q\mathbb{O}_{K_n} = \lambda_1\lambda_2\cdots\lambda_n$, a product of distinct prime ideals). If λ is any such prime ideal in $\mathbb{O}_{K_n}$, then $\mathbb{O}_{K_n}/\lambda \cong \mathbb{Z}/q\mathbb{Z}$. We have the congruence

$$(\vartheta(x))^q = (\psi_p(x) - \tilde{\psi}_p(x))^q \equiv \left(\frac{q}{p}\right)_{2n} (\psi_p(x^q) - \tilde{\psi}_p(x^q)) \pmod{\lambda}.$$

On the other hand, we also have

$$(\vartheta(x))^q = (\sqrt{\beta}\,\phi(x))^q \equiv \beta^{(q-1)/2}\sqrt{\beta}\,\phi(x^q) \pmod{\lambda}$$
$$\equiv \left(\frac{\beta}{\lambda}\right)(\psi_p(x^q) - \tilde{\psi}_p(x^q)) \pmod{\lambda}.$$

We will obtain the desired result from the congruence

$$\left(\frac{q}{p}\right)_{2n}(\psi_p(x^q) - \tilde{\psi}_p(x^q)) \equiv \left(\frac{\beta}{\lambda}\right)(\psi_p(x^q) - \tilde{\psi}_p(x^q)) \pmod{\lambda}$$

once we show that $\psi_p(X) \not\equiv \tilde{\psi}_p(X) \pmod{\lambda}$; note that if $\psi_p(X) \equiv \tilde{\psi}_p(X) \pmod{\lambda}$, then $\varphi_p(X) \equiv \psi(X)^2 \pmod{\lambda}$. Applying Kummer's theorem [Janusz 1996, Theorem 7.4], the polynomial $\Phi_p(X)$ factors in exactly the same way in

$$(\mathbb{Z}/q\mathbb{Z})[X] \cong (\mathbb{O}_{K_n}/\lambda)[X],$$

as $q\mathbb{Z}[\zeta_p]$ factors in $\mathbb{Z}[\zeta_p]$. However, the distinctness of the primes p and q implies that $q\mathbb{Z}[\zeta_p]$ does not ramify, giving a contradiction. Thus, we conclude that

$$\left(\frac{q}{p}\right)_{2n} \equiv \left(\frac{\beta}{\lambda}\right) \pmod{\lambda},$$

which reduces to an equality since the residue symbols only take on the values ± 1. □

While this reciprocity law may not appear to be rational, given the existence of the quadratic residue symbol, it can be identified with a Legendre symbol. Namely, the element β is a coset representative in

$$\mathbb{O}_{K_n}/\lambda \cong \mathbb{Z}/q\mathbb{Z},$$

and since $0, 1, \ldots, q-1$ represent distinct cosets in $\mathbb{O}_{K_n}/\lambda$, we have $\beta \equiv a \pmod{\lambda}$ for some unique element $a \in \{1, 2, \ldots, q-1\}$. Thus, we have

$$\left(\frac{\beta}{\lambda}\right) = \left(\frac{a}{\lambda}\right),$$

and since Theorem 1 is independent of the choice of prime λ above q, we may write

$$\left(\frac{\beta}{\lambda}\right) = \left(\frac{a}{q}\right).$$

In this capacity, Theorem 1 may be viewed as a rational reciprocity law.

We chose the polynomial-based proof given for Theorem 1 because it highlights the significance of Kummer's theorem, relating the factoring of minimal polynomials in function fields to that of prime ideals in number fields. We note that Theorem 1 can also be proved in an analogous way to Lemmermeyer's proof of the all-encompassing rational quartic reciprocity law in [Lemmermeyer 1994].

3. Generalizing Burde's law

Since Theorem 1 is a generalization of the all-encompassing rational quartic reciprocity law in [Lemmermeyer 1994], the rational quartic laws of Scholz [1934], Lehmer [1958; 1978] and Burde [1969] all follow by picking appropriate primitive elements for K_4. In this section, we show that Theorem 1 implies a generalization of Burde's law that differs from the known generalizations. Before giving the general case, we recall Lemmermeyer's proof [2000] of Burde's law for motivation.

Assume that $p \equiv q \equiv 1 \pmod 4$ are distinct primes, so we can write $p = a^2 + b^2$ and $q = A^2 + B^2$ with $2 \nmid aA$. We also assume that $(p/q) = 1$. A few simple consequences of these conditions that can be checked directly are

$$\left(\frac{A}{q}\right) = 1 \quad \text{and} \quad \left(\frac{2B}{q}\right) = 1.$$

Lemmermeyer argued that $K_4 = \mathbb{Q}(\sqrt{\beta_4})$, where

$$\beta_4 = pq + (b(A^2 - B^2) + 2aAB)\sqrt{p}.$$

Then we see that

$$\begin{aligned}
\left(\frac{\beta_4}{q}\right) \equiv \beta_4^{(q-1)/2} &\equiv (b(A^2 - B^2) + 2aAB)^{(q-1)/2} p^{(q-1)/4} \pmod q \\
&\equiv (-2bB^2 + 2aAB)^{(q-1)/2} \left(\frac{p}{q}\right)_4 \pmod q \\
&\equiv (-2B(bB - aA))^{(q-1)/2} \left(\frac{p}{q}\right)_4 \pmod q \\
&\equiv \left(\frac{-2B}{q}\right)\left(\frac{bB - aA}{q}\right)\left(\frac{p}{q}\right)_4 \pmod q \\
&\equiv \left(\frac{bB - aA}{q}\right)\left(\frac{p}{q}\right)_4 \pmod q.
\end{aligned}$$

Thus, from Theorem 1, we obtain Burde's law:

$$\left(\frac{p}{q}\right)_4 \left(\frac{q}{p}\right)_4 = \left(\frac{bB - aA}{q}\right).$$

Note that Burde's law is independent of the choices of signs of a, b, A, and B.

We now describe a primitive element for K_{2^t}, when $t \geq 2$, analogous to $\sqrt{\beta_4}$ used above for K_4.

Theorem 2. *Let $p \equiv q \equiv 1 \pmod{2^t}$ be distinct primes with $t \geq 2$ such that $p = a^2 + b^2$ and $q = A^2 + B^2$ with $2 \nmid aA$. If $\beta_4 = pq + (b(A^2 - B^2) + 2aAB)\sqrt{p}$, then a primitive element for K_{2^t} can be defined recursively for $t > 2$ by*

$$\beta_{2^t} = \left(q\sqrt{p} + (b(A^2 - B^2) + 2aAB)\right)\sqrt{\beta_{2^{t-1}}},$$

with $K_{2^t} = \mathbb{Q}(\sqrt{\beta_{2^t}})$.

Proof. Our proof proceeds by using (weak) induction on $t \geq 2$ following Lemmermeyer's approach [Lemmermeyer 1994] in the quartic case (and as our starting point when $t = 2$). Assume that the theorem holds for the 2^{t-1} case with $K_{2^{t-1}} = \mathbb{Q}(\sqrt{\beta_{2^{t-1}}})$ and let

$$\alpha_{2^t} = q\sqrt{p}\sqrt{\beta_{2^{t-1}}}, \quad \gamma = (b(A^2 - B^2) + 2aAB), \quad \delta = (a(A^2 - B^2) - 2bAB).$$

It is easily checked that α_{2^t}, γ, and δ are pairwise relatively prime and that

$$\alpha_{2^t}^2 = \beta_{2^{t-1}}(\gamma^2 + \delta^2).$$

From the identity

$$2\big(\alpha_{2^t} + \gamma\sqrt{\beta_{2^{t-1}}}\big)\big(\alpha_{2^t} + \delta\sqrt{\beta_{2^{t-1}}}\big) = \big(\alpha_{2^t} + \gamma\sqrt{\beta_{2^{t-1}}} + \delta\sqrt{\beta_{2^{t-1}}}\big)^2,$$

we see that

$$K_{2^t} := \mathbb{Q}\Big(\sqrt{\alpha_{2^t} + \gamma\sqrt{\beta_{2^{t-1}}}}\Big) = \mathbb{Q}\Big(\sqrt{2(\alpha + \delta\sqrt{\beta_{2^{t-1}}})}\Big).$$

Thus, the only primes that can possibly ramify in $K_{2^t}/K_{2^{t-1}}$ are 2 and any common divisors of

$$\alpha_{2^t}^2 - \beta_{2^{t-1}}\gamma^2 = \beta_{2^{t-1}}\delta^2 \quad \text{and} \quad \alpha_{2^t}^2 - \beta_{2^{t-1}}\delta^2 = \beta_{2^{t-1}}\gamma^2.$$

Since δ and γ are relatively prime, the only odd primes that can ramify are divisors of $\beta_{2^{t-1}}$. However, any such prime would have to have ramified in $\mathbb{Q}\big(\sqrt{\beta_{2^{t-1}}}\big)$ and by our inductive hypothesis, only p ramified there. Thus, p is the only odd prime that ramifies in $K_{2^t}/K_{2^{t-1}}$.

Finally, we must argue that 2 does not ramify. Lemmermeyer [1994] showed the case $t = 2$, that is, $\beta_4 \equiv 1 \pmod 4$. As our inductive hypothesis, we assume that $\beta_{2^{t-1}} \equiv 1 \pmod 4$. Then the congruences

$$\sqrt{\beta_{2^{t-1}}} \equiv \pm 1 \pmod 4, \quad \sqrt{p} \equiv \pm 1 \pmod 4, \quad q \equiv 1 \pmod 4$$

and the fact that γ is even show that $\beta_{2^t} \equiv \sqrt{\beta_{2^{t-1}}}(q\sqrt{p} + \gamma) \equiv \pm 1 \pmod 4$. By Stickelberger's discriminant relation [Ribenboim 2001, Section 6.3], the discriminant of an algebraic number field is $0, 1 \pmod 4$. Thus, $\beta_{2^t} \equiv 1 \pmod 4$ and we conclude that 2 does not ramify in $K_{2^t}/K_{2^{t-1}}$. Since p is the only prime that ramifies in the abelian Galois extension $K_{2^t}/\mathbb{Q}$, K_{2^t} is the unique subfield of $\mathbb{Q}(\zeta_p)$ of degree 2^t over $\mathbb{Q}$ by the theorem of Kronecker and Weber [Ribenboim 2001, Section 15.1]. □

Using the reciprocity law given in Theorem 1 with the choice of primitive element for K_{2^t} given in Theorem 2, we obtain the following 2^t-th generalization of Burde's law, which is also independent of the choices of signs of a, b, A, and B.

Theorem 3. *Let $p \equiv q \equiv 1 \pmod{2^t}$ be distinct primes with $t \geq 2$ such that*

$$p = a^2 + b^2 \quad \textit{and} \quad q = A^2 + B^2,$$

with $2 \nmid aA$. If

$$\left(\frac{p}{q}\right)_{2^{t-1}} = \left(\frac{q}{p}\right)_{2^{t-1}} = 1,$$

then

$$\left(\frac{p}{q}\right)_{2^t}\left(\frac{q}{p}\right)_{2^t} = \left(\frac{2B(bB - aA)}{q}\right)_{2^{t-1}}.$$

Proof. Once again, we use an inductive argument with Lemmermeyer's proof of Burde's law as a starting point. With regard to Theorem 1, assuming that Theorem 3 is true for the $t-1$ case is equivalent to assuming that

$$\left(\frac{\beta_{2^{t-1}}}{q}\right) = \left(\frac{2B(bB - aA)}{q}\right)_{2^{t-2}}\left(\frac{p}{q}\right)_{2^{t-1}}.$$

Letting $\left(\frac{p}{q}\right)_{2^{t-1}} = \left(\frac{q}{p}\right)_{2^{t-1}} = 1$, we then obtain, for $t > 2$,

$$\begin{aligned}\left(\frac{q}{p}\right)_{2^t} = \left(\frac{\beta_{2^t}}{q}\right) &\equiv \beta_{2^t}^{(q-1)/2} \equiv \beta_{2^{t-1}}^{(q-1)/4}(b(A^2 - B^2) + 2aAB)^{(q-1)/2} \pmod{q} \\ &\equiv \left(\frac{2B(bB - aA)}{q}\right)_{2^{t-1}}\left(\frac{p}{q}\right)_{2^t}\left(\frac{2B(bB - aA)}{q}\right) \pmod{q} \\ &\equiv \left(\frac{2B(bB - aA)}{q}\right)_{2^{t-1}}\left(\frac{p}{q}\right)_{2^t} \pmod{q}.\end{aligned}$$

Since all of the rational residue symbols take on only the values ± 1, we may drop the congruence and conclude the statement of Theorem 3. □

Perhaps the other known generalizations of Burde's law also follow as consequences of Theorem 1. At this time, we have not been able to find suitable primitive elements to prove such implications.

References

[Budden et al. 2007] M. Budden, J. Eisenmenger, and J. Kish, "A generalization of Scholz's reciprocity law", *J. Théor. Nombres Bordeaux* **19**:3 (2007), 583–594. MR 2009b:11004

[Burde 1969] K. Burde, "Ein rationales biquadratisches Reziprozitätsgesetz", *J. Reine Angew. Math.* **235** (1969), 175–184. MR 39 #2694

[Evans 1981] R. J. Evans, "Rational reciprocity laws", *Acta Arith.* **39**:3 (1981), 281–294. MR 83h:10006 MR 83h:10006 Zbl 0472.10006

[Evans 1989] R. Evans, "Residuacity of primes", *Rocky Mountain J. Math.* **19**:4 (1989), 1069–1081. MR 90m:11008 Zbl 0699.10012

[Gallian 2010] J. Gallian, *Contemporary Abstract Algebra*, 7th ed., Brooks Cole, Belmont, CA, 2010.

[Janusz 1996] G. J. Janusz, *Algebraic number fields*, 2nd ed., Graduate Studies in Mathematics **7**, American Mathematical Society, Providence, RI, 1996. MR 96j:11137 Zbl 0854.11001

[Lehmer 1958] E. Lehmer, "Criteria for cubic and quartic residuacity", *Mathematika* **5** (1958), 20–29. MR 20 #1668 Zbl 0102.28002

[Lehmer 1978] E. Lehmer, "Rational reciprocity laws", *Amer. Math. Monthly* **85**:6 (1978), 467–472. MR 58 #16482 Zbl 0383.10003

[Lemmermeyer 1994] F. Lemmermeyer, "Rational quartic reciprocity", *Acta Arith.* **67**:4 (1994), 387–390. MR 95m:11010 Zbl 0833.11049

[Lemmermeyer 2000] F. Lemmermeyer, *Reciprocity laws:From Euler to Eisenstein*, Springer, Berlin, 2000. MR 2001i:11009 Zbl 0949.11002

[Leonard and Williams 1977] P. A. Leonard and K. S. Williams, "A rational sixteenth power reciprocity law", *Acta Arith.* **33**:4 (1977), 365–377. MR 57 #219 Zbl 0363.10003

[Ribenboim 2001] P. Ribenboim, *Classical theory of algebraic numbers*, Universitext, Springer, New York, 2001. MR 2002e:11001 Zbl 1082.11065

[Scholz 1934] A. Scholz, "Über die Lösbarkeit der Gleichung $t^2 - Du^2 = -4$", *Math. Z.* **39** (1934), 95–111.

[Williams 1976] K. S. Williams, "A rational octic reciprocity law", *Pacific J. Math.* **63**:2 (1976), 563–570. MR 54 #2568 Zbl 0311.10004

[Williams et al. 1985] K. S. Williams, K. Hardy, and C. Friesen, "On the evaluation of the Legendre symbol $((A + B\sqrt{m})/p)$", *Acta Arith.* **45**:3 (1985), 255–272. MR 87b:11006 Zbl 0524.10002

[Wu 1975] P. Wu, "A rational reciprocity law", Ph.D. thesis, University of Southern California, Los Angeles, 1975.

Received: 2009-04-27 Accepted: 2010-09-20

mrbudden@email.wcu.edu *Department of Mathematics and Computer Science, Western Carolina University, Cullowhee, NC 28723, United States*

ac0428@students.armstrong.edu *Department of Mathematics, Armstrong Atlantic State University, 11935 Abercorn St., Savannah, GA 31419, United States*

ke3203@students.armstrong.edu *Department of Mathematics, Armstrong Atlantic State University, 11935 Abercorn St., Savannah, GA 31419, United States*

ss7965@students.armstrong.edu *Department of Mathematics, Armstrong Atlantic State University, 11935 Abercorn St., Savannah, GA 31419, United States*

INVOLVE 3:3(2010)

Coexistence of stable ECM solutions in the Lang–Kobayashi system

Ericka Mochan, C. Davis Buenger and Tamas Wiandt

(Communicated by John Baxley)

The Lang–Kobayashi system of delay differential equations describes the behavior of the complex electric field $\mathscr{E}$ and inversion N inside an external cavity semiconductor laser. This system has a family of special periodic solutions known as external cavity modes (ECMs). It is well known that these ECM solutions appear through saddle-node bifurcations, then lose stability through a Hopf bifurcation before new ECM solutions are born through a secondary saddle-node bifurcation. Employing analytical and numerical techniques, we show that for certain short external cavity lasers the loss of stability happens only after the secondary saddle-node bifurcations, which means that stable ECM solutions can coexist in these systems. We also investigate the basins of these ECM attractors.

1. Introduction

Today nonlinear delay differential equations (NDDEs) are used extensively in many fields of science and engineering. Disciplines such as population dynamics, epidemiology, financial mathematics, and optoelectronics, to name a few, use NDDEs in their modeling efforts. In most cases the model equations have very simple functional forms, yet this apparent simplicity is deceiving. They display unusually rich and complex dynamics, which primarily is a result of the high dimensionality that the time-delayed terms introduce [Hale and Verduyn Lunel 1993; Driver 1977].

Our focus is on equations modeling the behavior of external cavity semiconductor lasers. Semiconductor lasers offer many advantages not only due to their compact size but also because of their enormous application in various fields, particularly in optical data recording and optical fiber communications.

Optical feedback is inevitable in virtually all realistic applications, which can be due to, for instance, reflections from fiber facets when radiation is coupled into a fiber. From the standpoint of dynamics, an optical feedback introduces a time delay to the reinjected field which in turn makes the phase space dimension of the

MSC2000: 37G35, 37M20, 78A60.

Keywords: delay differential equations, bifurcations, Lang–Kobayashi equations.

underlying dynamical system infinite. The high dimensionality renders the analysis and understanding of external cavity lasers an extremely challenging problem from the dynamical systems point of view. As a result, our fundamental understanding about the bifurcation mechanisms leading to chaotic responses is still lacking [Davidchack et al. 2000; 2001; Erneux et al. 2000].

The performance of semiconductor lasers can be degraded significantly when the feedback is at moderate or high levels.

In general, the study of a nonlinear system often begins with an analysis of certain types of stationary solutions or fixed points. The method of study of the Lang–Kobayashi equation is similar in this respect: first we identify specific solutions, then we attempt to interpret the behavior of the model at different parameter values in terms of the location and stability properties of these specific solutions.

Lang and Kobayashi [1980] formulated a model consisting of two delay differential equations for the complex electrical field $\mathscr{E}$ and the carrier number N (see also [Alsing et al. 1996; Heil et al. 2001; 2003]). Numerical simulations have shown that these equations correctly describe the experimentally observed dominant effects. The equations are given by

$$\frac{d\mathscr{E}}{dt} = (1+i\alpha)N\mathscr{E} + \eta e^{-i\omega_0\tau}\mathscr{E}(t-\tau), \tag{1}$$

$$T\frac{dN}{dt} = P - N - (1+2N)\,|\mathscr{E}|^2 \tag{2}$$

where $\mathscr{E} = E_x(t) + iE_y(t)$. The physical interpretation of $\mathscr{E}$ is the complex electric field of the laser, and $N(t)$ is the carrier number density of the laser. The parameters involved are α, the line-width enhancement factor; η, the feedback strength; τ, the external cavity round-trip time; ω_0, the angular frequency; T, the ratio of carrier lifetime to photon lifetime; and P, the dimensionless pump current. The physically meaningful values we use in our investigation are $\alpha = 5,\ \tau = 5,\ T = 1710,$ $P = 1.155$. These values were also used in [Heil et al. 2003]. We will make the usual assumption $\omega_0 = -\arctan\alpha/\tau$ to simplify our computations. Our bifurcation parameter will be η.

By setting $\mathscr{E} = E_x(t) + iE_y(t)$, the equations can be expressed as

$$\dot{E}_x(t) = NE_x - \alpha NE_y + \eta(\cos(\omega_0\tau)E_x(t-\tau) + \sin(\omega_o\tau)E_y(t-\tau)), \tag{3}$$

$$\dot{E}_y(t) = \alpha NE_x + NE_y + \eta(-\sin(\omega_o\tau)E_x(t-\tau) + \cos(\omega_0\tau)E_y(t-\tau)), \tag{4}$$

$$\dot{N} = \frac{1}{T}(P - N - (1+2N)(E_x^2 + E_y^2)). \tag{5}$$

We will use this form in our numerical analysis with Matlab and the Matlab package DDE-BIFTOOL [Engelborghs et al. 2002].

2. External cavity modes

Solutions to the system vary depending on the chosen values of the parameters. A certain type of solution is an external cavity mode, or ECM. The ECM is a specific solution with a constant carrier number density and constant light intensity [Rottschäfer and Krauskopf 2007]. The ECM is typically of the form

$$\mathscr{E} = E_s e^{i\phi_s t}, \quad N = N_s,$$

where E_s, ϕ_s, and N_s are constants. This can be substituted into the complex-form equations to solve for the variable ϕ_s in terms of the original parameters:

$$E_s i\phi_s e^{i\phi_s t} = (1+i\alpha)N_s E_s e^{i\phi_s t} + \eta e^{-i\omega_0\tau} E_s e^{i\phi_s(t-\tau)}, \tag{6}$$

$$0 = P - N_s - (1+2N_s)E_s^2. \tag{7}$$

Dividing (6) by $e^{i\phi_s t}$ gives us

$$E_s i\phi_s = (1+i\alpha)N_s E_s + \eta E_s e^{-i(\omega_0\tau+\phi_s\tau)}.$$

Assuming $E_s \neq 0$, the equation can be divided by E_s to find

$$i\phi_s = (1+i\alpha)N_s + \eta e^{-i(\omega_0\tau+\phi_s\tau)}. \tag{8}$$

Comparing real and imaginary parts of (7) and (8), we obtain

$$0 = N_s + \eta\cos(\tau(\omega_0+\phi_s)), \tag{9}$$

$$\phi_s = \alpha N_s - \eta\sin(\tau(\omega_0+\phi_s)), \tag{10}$$

$$0 = P - N_s - (1+2N_s)E_s^2. \tag{11}$$

To find ϕ_s, we use (9) and (10) to eliminate N_s and get

$$-\phi_s = \alpha\eta\cos(\tau(\omega_0+\phi_s)) + \eta\sin(\tau(\omega_0+\phi_s)).$$

Then setting $\beta = \arctan\alpha$, we have

$$\tan\beta = \alpha, \qquad \sin\beta = \frac{\alpha}{\sqrt{\alpha^2+1}}, \qquad \cos\beta = \frac{1}{\sqrt{\alpha^2+1}}.$$

Using the trigonometric identity for $\sin(x+y)$ we obtain

$$-\phi_s = \eta\sqrt{\alpha^2+1}\,\sin(\arctan\alpha + \tau(\omega_0+\phi_s)).$$

Since we are assuming $\tau\omega_0 = -\arctan\alpha$,

$$-\phi_s = \eta\sqrt{\alpha^2+1}\sin(\tau\phi_s).$$

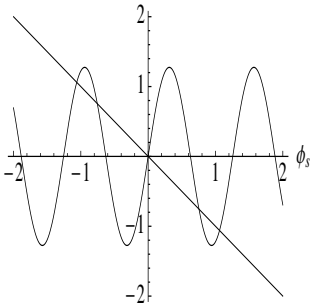


Figure 1. Graph of the two sides of (12).

The final equations are therefore

$$-\phi_s = \eta\sqrt{\alpha^2+1}\sin(\tau\phi_s), \tag{12}$$

$$N_s = -\eta\cos(\tau(\omega_0+\phi_s)), \tag{13}$$

$$E_s = \sqrt{\frac{P-N_s}{1+2N_s}}. \tag{14}$$

Since (12) is transcendental, a closed form solution cannot be obtained, so numerical solutions will be found. An example plot of the two sides of (12) is given in Figure 1, using our values $\alpha = 5$, $\tau = 5$, and $\eta = 0.25$. As we can see, both sides of (12) are odd functions, so for any ϕ_s solution, $-\phi_s$ is also a solution. Also, $\phi_s = 0$ is always a solution of (12), which gives us a family of equilibrium points situated on a circle in the phase-space. We will not consider the behavior of these degenerate ECMs in this paper (for different values of the bifurcation parameter η, the stability of these equilibrium points changes as well).

3. Bifurcations

ECM solutions appear as a result of saddle-node bifurcations and disappear with the occurrence of Hopf bifurcations. Changing η, the bifurcation parameter, will change the amplitude of the right side of (12). This will change the number of solutions. First, we only have $\phi_s = 0$, then by changing the amplitude, we have two new solutions at tangency points, and finally we have four solutions at four distinct intersection points. This bifurcation is demonstrated below in Figure 2. It is clear that (12) always has $\phi_s = 0$ as a solution. For the other intersections, we use the fact that at the bifurcation we have a tangency. At the point of tangency, the derivatives of both sides of (12) are equal:

$$-1 = \eta\sqrt{\alpha^2+1}\cos(\tau\phi_s)\tau.$$

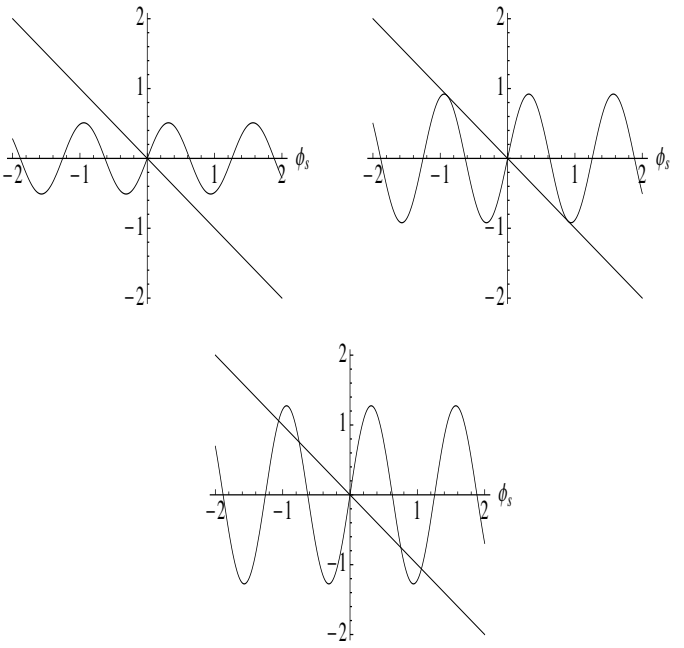


Figure 2. Example of bifurcation by changing η. Clockwise from top left: $\eta = 0.1, 0.1806, 0.25$.

So, the equations for the system at the tangency point are

$$\eta = \frac{-1}{\tau} \cdot \frac{1}{\sqrt{1+\alpha^2}} \cdot \frac{1}{\cos(\phi_s \tau)},$$
$$-\phi_s = \frac{-1}{\tau} \cdot \frac{1}{\sqrt{1+\alpha^2}} \cdot \frac{1}{\cos(\phi_s \tau)} \cdot \sqrt{1+\alpha^2} \cdot \sin(\phi_s \tau).$$

This means that at the tangency,

$$\phi_s \tau = \tan(\phi_s \tau).$$

Considering the graphs of x and $\tan x$ we can see that the solutions of the previous equation are on the intervals $((2n-1)\pi/2, (2n+1)\pi/2)$. We are only interested in the solutions on the intervals $((4n+1)\pi/2, (4n+3)\pi/2)$, because the solutions ϕ_s on the other intervals give us a negative η value. Also, asymptotically the solutions of this equation are $\phi_s \tau \sim (4n+3)\pi/2$.

For example, for our values of $\tau = 5$, $\alpha = 5$, $P = 1.155$, $T = 1710$, the first saddle-node bifurcation occurs at $\eta \approx 0.1806$, and the second saddle-node bifurcation is at $\eta \approx 0.4295$.

4. Stability of ECMs

Generally, in the case of ordinary differential equations, when a saddle-node bifurcation occurs, one of the equilibrium points created is stable and the other is unstable. In our case, the saddle-node bifurcation creates four ECM solutions (two pairs, one pair for the negative ϕ_s values and one pair for the positive ϕ_s values). In one of these pairs, both ECM solutions are unstable (when ϕ_s is positive) and in the other pair, one ECM solution is stable and the other is unstable.

As an illustration, in Figures 3 and 4 we plot two solutions for the value $\eta = 0.4$ with history $\mathscr{E} = E_s e^{i\phi_s t}$, $N = N_s$, where ϕ_s, N_s and E_s are obtained from (12), (13), and (14). The values are $\phi_s \approx -1.1382$, $N_s \approx -0.2837$, $E_s \approx 1.8250$ and $\phi_s \approx -0.6982$, $N_s \approx -0.0606$, $E_s \approx 1.760$. As the figure shows, one of the ECMs is stable and the other one is unstable.

We used the Matlab function `dde23` to create the illustration below.

The Matlab package DDE-BIFTOOL was used to analyze the stability of equilibrium points and periodic solutions. Using this package, we calculate a branch

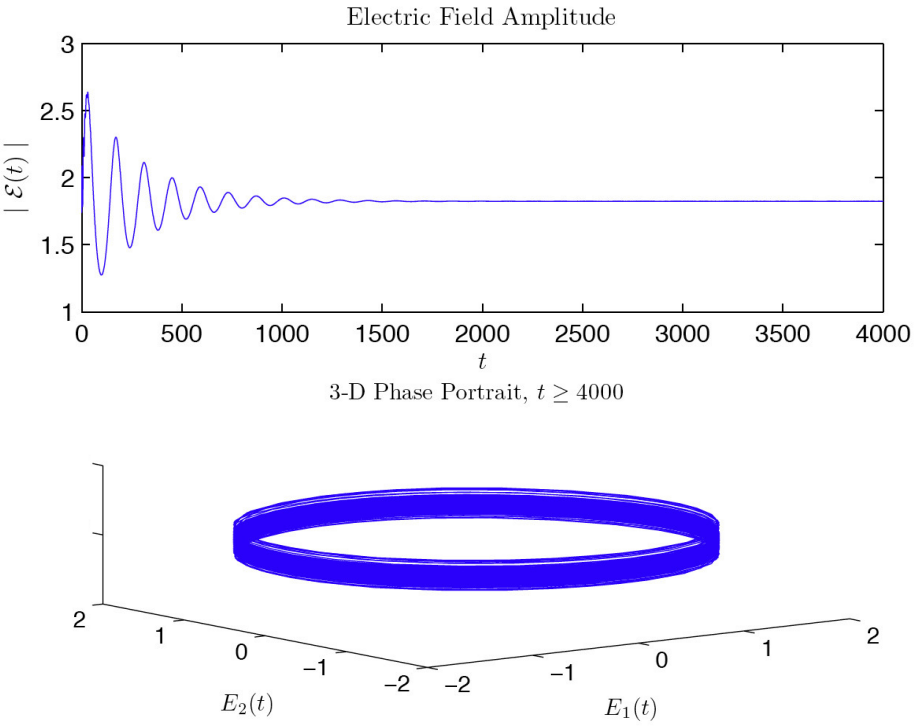


Figure 3. Stable ECM solution. The vertical coordinate in the three-dimensional graph, $N(t)$, is approximately -0.2837.

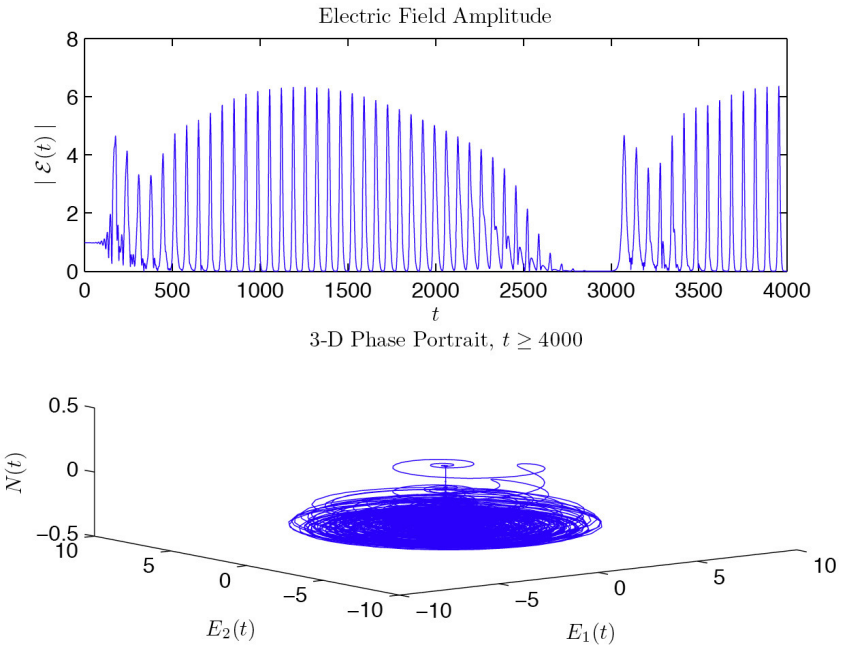


Figure 4. Unstable ECM solutions.

of ECM solutions over a range of η values. The branch plot in Figure 5 shows the amplitude of ECM solutions versus the feedback parameter η (each point on this figure represents an ECM).

On Figure 6, we show for different values of η the corresponding ECM solutions on the branch figure.

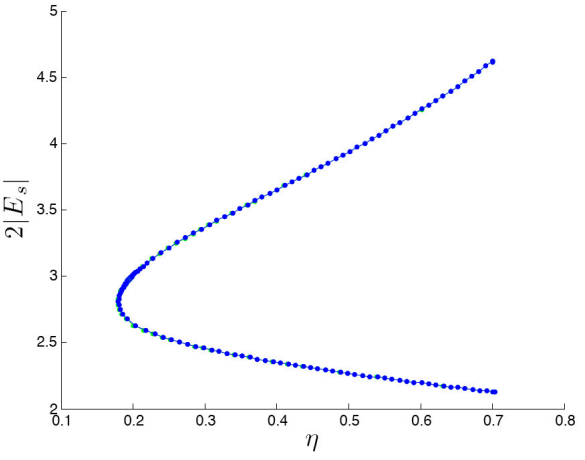


Figure 5. Branch plot from Matlab.

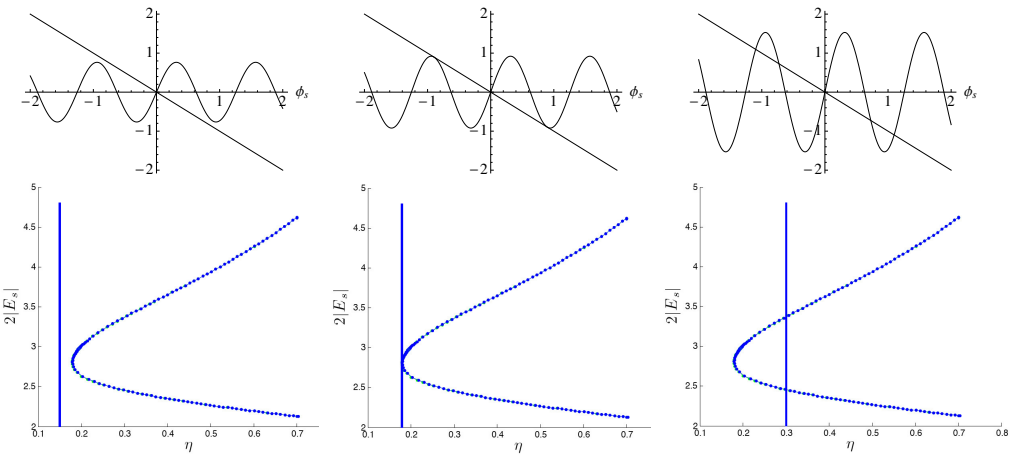


Figure 6. ECM solutions for different values of η.

Floquet multipliers are also calculated with DDE-BIFTOOL and are used to determine the stability of our ECM solutions. In order to be stable, Floquet multipliers must have an absolute value less than 1. (There is always one Floquet multiplier equal to 1, but that does not affect the stability of the periodic solution.)

The Floquet multipliers are inside the unit circle on Figure 7 (left), which proves the stability of that ECM solution. In Figure 7 (right), some of the Floquet multipliers are outside the unit circle, so the corresponding periodic solution (ECM) is unstable. This matches our numerical observations by `dde23` presented earlier in this section.

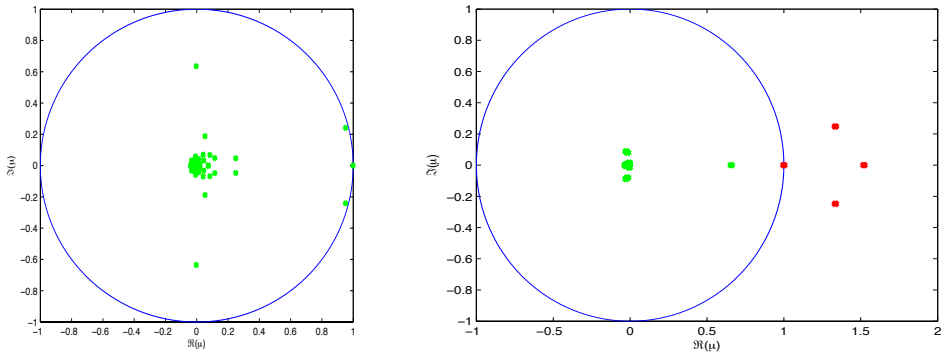


Figure 7. Floquet multipliers of stable solutions (left) and unstable ones (right).

5. Coexistence of stable ECMs and basins of attraction

Typically the leftmost ϕ_s value corresponds to the stable ECM solution and the other ECM solutions are unstable. It was observed earlier that after ECM solutions are created by saddle-node bifurcations, the stable ECM solution loses stability through a Hopf bifurcation for a slightly higher η value, and then another stable ECM will emerge through a new saddle-node bifurcation. On Figure 8, we illustrate the loss of stability through a Hopf bifurcation.

For short external cavity semiconductor lasers, there is a possibility of coexistence of two stable ECM solutions. For the $\alpha = 5,\ \tau = 5$ case, we find that the Hopf bifurcation, through which the primary ECM loses stability, occurs only after the secondary ECM is born. This creates two simultaneous stable ECM solutions. This coexistence of stable ECMs is maintained for τ values up to $\tau \approx 35$.

Using the calculated branch of the ECMs, the stability of these ECMs was determined. Figure 9 plots the absolute value of the Floquet multipliers as a function of η. Figure 9 (left) shows Floquet multipliers for the branch emerging from the primary bifurcation point, and Figure 9 (right) shows that of the branch emerging from the secondary bifurcation point. The graphs provide a rough estimate of the η value where the ECMs lose stability. Analysis of this figure reveals the coexistence of stable ECMs on the approximate range $0.43 < \eta < 0.53$.

The coexistence of two stable ECMs creates a partition in the history function space between solutions that converge to the first ECM and those that converge to the second. Of course, this function space is an infinite dimensional space, so we will consider a three dimensional subspace consisting of periodic solutions in the form $\mathscr{E}(t) = Ee^{i\phi t}$, $N(t) = N$. Figures 10–14 demonstrate the basins for the two stable ECMs for various η values. White dots indicate an initial condition function for which the solution converges to the ECM from the primary bifurcation, and

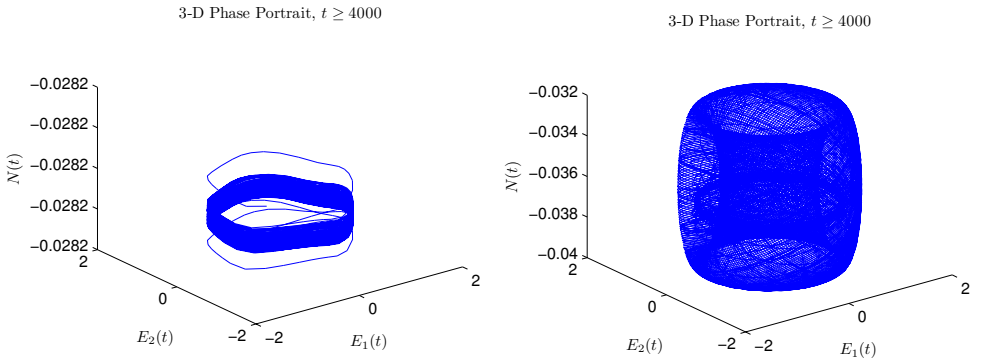


Figure 8. Left: primary ECM; right: Hopf bifurcation.

black dots indicate the initial condition functions for which the solution converges to the ECM from the secondary bifurcation. On each figure, the six subfigures correspond to the specified N value, the ϕ and E values on every subfigure correspond to the range specified on the first subfigure, divided evenly between the given values.

As these figures show, the basin of the secondary ECM attractor is growing as η increases. Accordingly, the basin of the primary ECM attractor is contracting before this ECM loses stability through the above-mentioned Hopf bifurcation at around $\eta = 0.53$.

We demonstrated that for certain short external cavity semiconductor lasers, the coexistence of stable ECM solutions is possible. Computations indicate that

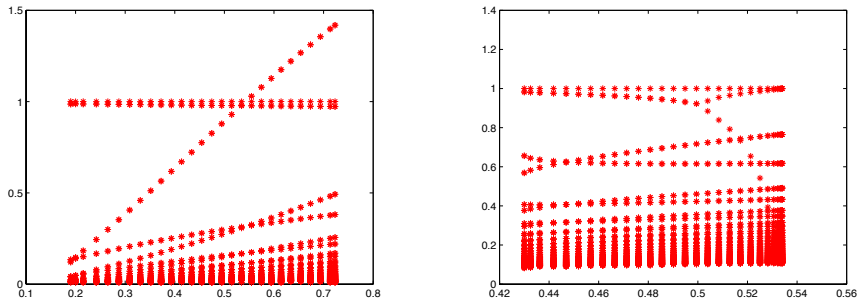


Figure 9. Floquet multipliers at the primary and secondary bifurcations.

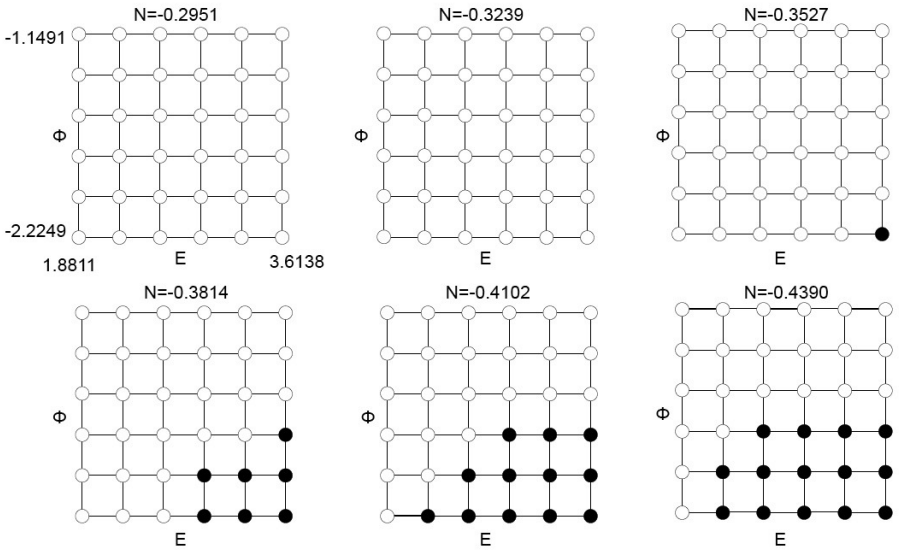


Figure 10. The case $\eta = 0.44$.

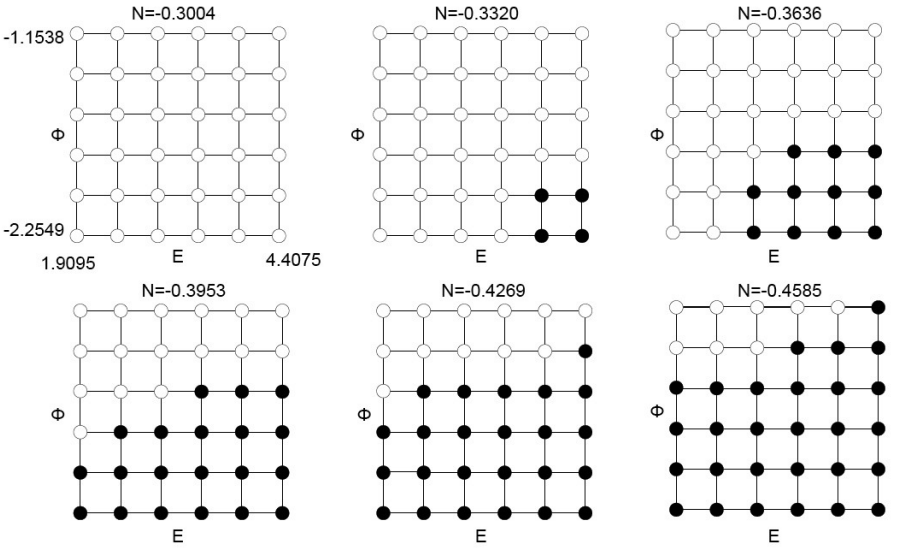


Figure 11. The case $\eta = 0.46$.

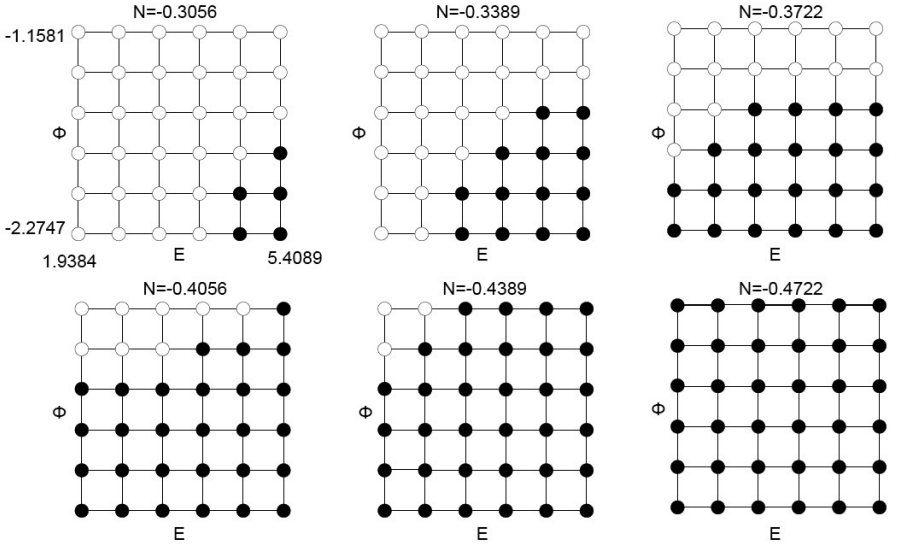


Figure 12. The case $\eta = 0.48$.

this coexistence of the stable primary and secondary ECM solutions disappear at around $\tau \approx 35$ (for the previously specified α, P, T values). This means that for short external cavities there is a range of the feedback parameter η where the laser can operate in two different modes.

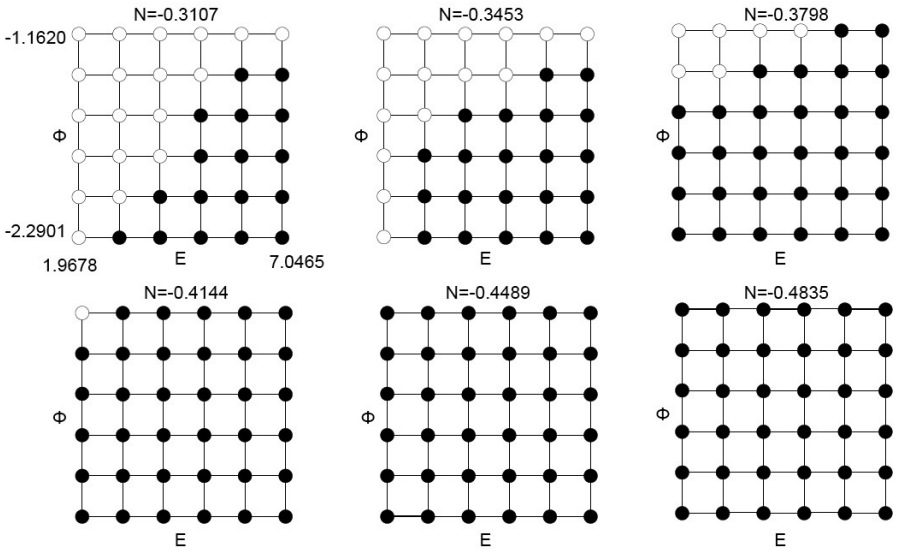


Figure 13. The case $\eta = 0.50$.

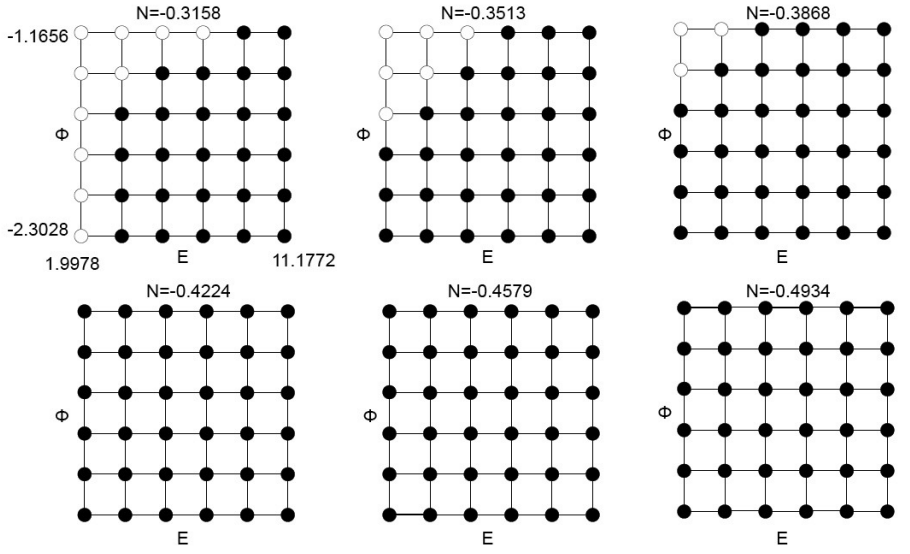


Figure 14. The case $\eta = 0.52$.

References

[Alsing et al. 1996] P. M. Alsing, V. Kovanis, A. Gavrielides, and T. Erneux, "Lang and Kobayashi phase equation", *Phys. Rev. A* **53** (1996), 4429–4434.

[Davidchack et al. 2000] R. L. Davidchack, Y. C. Lai, A. Gavrielides, and V. Kovanis, "Dynamical origins of low frequency fluctuations in external cavity semiconductor lasers", *Phys. Lett. A* **267** (2000), 350–356.

[Davidchack et al. 2001] R. L. Davidchack, Y. C. Lai, A. Gavrielides, and V. Kovanis, "Regular dynamics of low frequency fluctuations in external cavity semiconductor lasers", *Phys. Rev. E* **63** (2001), Art. 056256.

[Driver 1977] R. D. Driver, *Ordinary and delay differential equations*, Applied Math. Sciences **20**, Springer, New York, 1977. MR 57 #16897 Zbl 0374.34001

[Engelborghs et al. 2002] K. Engelborghs, T. Luzyanina, and D. Roose, "Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL", *ACM Trans. Math. Software* **28**:1 (2002), 1–21. MR 1918642 Zbl 1070.65556

[Erneux et al. 2000] T. Erneux, F. Rogister, A. Gavrielides, and V. Kovanis, "Bifurcation to mixed external cavity mode solutions for semiconductor lasers subject to optical feedback", *Optics Comm.* **183** (2000), 467–477.

[Hale and Verduyn Lunel 1993] J. K. Hale and S. M. Verduyn Lunel, *Introduction to functional-differential equations*, Applied Math. Sciences **99**, Springer, New York, 1993. MR 94m:34169 Zbl 0787.34002

[Heil et al. 2001] T. Heil, I. Fischer, W. Elsässer, and A. Gavrielides, "Dynamics of semiconductor lasers subject to delayed optical feedback: The short cavity regime", *Phys. Rev. Lett.* **87** (2001), Art. 243901.

[Heil et al. 2003] T. Heil, I. Fischer, W. Elsäßer, B. Krauskopf, A. Gavrielides, and K. Green, "Delay dynamics of semiconductor lasers with short external cavities: bifurcation scenarios and mechanisms", *Phys. Rev. E* (3) **67**:6 (2003), Art. 066214. MR 1995899

[Lang and Kobayashi 1980] R. Lang and K. Kobayashi, "External optical feedback effects on semiconductor laser properties", *IEEE J. Quantum Electron.* **16** (1980), 347–355.

[Rottschäfer and Krauskopf 2007] V. Rottschäfer and B. Krauskopf, "The ECM-backbone of the Lang–Kobayashi equations: a geometric picture", *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **17**:5 (2007), 1575–1588. MR 2008e:78022

Received: 2009-09-15 Revised: 2010-08-06 Accepted: 2010-08-19

em284475@wnec.edu *Department of Biomedical Engineering, Western New England College, Springfield, MA 01119, United States*

buenger@math.ohio-state.edu *Department of Mathematics, The Ohio State University, Columbus, OH 43201, United States*

tiwsma@rit.edu *School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623, United States*

INVOLVE 3:3(2010)

A complex finite calculus

Joseph Seaborn and Philip Mummert

(Communicated by Johnny Henderson)

We explore a complex extension of finite calculus on the integer lattice of the complex plane. $f:\mathbb{Z}[i]\to\mathbb{C}$ satisfies the discretized Cauchy–Riemann equations at z if $\mathrm{Re}(f(z+1)-f(z))=\mathrm{Im}(f(z+i)-f(z))$ and $\mathrm{Re}(f(z+i)-f(z))=-\mathrm{Im}(f(z+1)-f(z))$. From this principle arise notions of the discrete path integral, Cauchy's theorem, the exponential function, discrete analyticity, and falling power series.

1. Introduction

The theory of finite (or discrete) calculus, that is, finite differences, has been well established. In addition, a unified theory of time scales has been formulated that encompasses both continuous and discrete calculus (for real variables) [Bohner and Peterson 2001]. The subject of complex analysis builds a continuous calculus on the complex plane. A remaining, natural question is what can we say about finite calculus on the complex plane? There are multiple approaches to addressing this question, and unbeknownst to the authors until after this work was completed, the question has been explored before under the monikers of *discrete analytic functions*, *preholomorphic functions*, and *monodiffric functions of the first kind* [Duffin 1956; Ferrand 1944; Isaacs 1941; 1952; Kiselman 2005; Mercat 2001]. Consequently, we do not claim mathematical originality for any of these results; we only hope to present these ideas in a fresh context. The reader is hereby warned that some familiar terms and theorem names will be used throughout this paper with a new meaning derived from the discretized context. To avoid confusion, invocations of these terms in their standard usage will be designated as *classical*.

MSC2000: 30G25, 39A12.

Keywords: complex analysis, discrete analytic, finite calculus, finite differences, monodiffric, preholomorphic, Gaussian integers, integer lattice, discrete.

2. Definitions

Let $\mathbb{Z}[i] = \{x + iy : x \in \mathbb{Z}, y \in \mathbb{Z}\}$ denote the integer lattice in the complex plane. Let $f(z) = f(x, y) = u(x, y) + iv(x, y) : \Omega \to \mathbb{C}$, where Ω is a subset of $\mathbb{Z}[i]$. The partial derivative of u with respect to x, $\Delta_x u(x, y)$, can be calculated by a finite difference as

$$\Delta_x u(x, y) = u(x + 1, y) - u(x, y)$$

or more simply as $\Delta_x u(z) = u(z + 1) - u(z)$. Similarly,

$$\Delta_y u(x, y) = u(x, y + 1) - u(x, y)$$

and again,

$$\Delta_y u(z) = u(z + i) - u(z).$$

This allows for the natural definition of

$$\Delta_x f = \Delta_x u + i\,\Delta_x v \quad \text{and} \quad \Delta_y f = \Delta_y u + i\,\Delta_y v.$$

Note that $\Delta_x u(z)$ is defined at $\{z \in \Omega : z + 1 \in \Omega\}$ and $\Delta_y u(z)$ is defined at $\{z \in \Omega : z + i \in \Omega\}$. We have the following lemma for mixed partials:

Lemma 2.1. *If f is defined on a set Ω, then on $\{z \in \Omega : z+1, z+i, z+1+i \in \Omega\}$ we have*

$$\Delta_{xy} f(z) = \Delta_{yx} f(z).$$

Proof.
$$\begin{aligned}\Delta_{xy} f(z) &= \Delta_x(f(z + i) - f(z)) \\ &= f(z + i + 1) - f(z + 1) - f(z + i) + f(z) \\ &= f(z + 1 + i) - f(z + i) - f(z + 1) + f(z) \\ &= \Delta_y(f(z + 1) - f(z)) = \Delta_{yx} f(z).\end{aligned}$$ □

Definition 2.2. The discrete function f is *holomorphic* at z if it satisfies the discrete Cauchy–Riemann equations at z:

$$\Delta_x u(z) = \Delta_y v(z) \text{ and } \Delta_y u(z) = -\Delta_x v(z).$$

Definition 2.3. The *partial derivative* of f with respect to z is

$$\Delta f = \frac{\Delta_x f - i\,\Delta_y f}{2} = \frac{f(z + 1) - f(z) - i(f(z + i) - f(z))}{2}$$

and with respect to $\bar{z}$ is

$$\bar{\Delta} f = \frac{\Delta_x f + i\,\Delta_y f}{2} = \frac{f(z + 1) - f(z) + i(f(z + i) - f(z))}{2}.$$

All partial derivative operators — Δ, $\bar{\Delta}$, Δ_x, Δ_y — are linear operators. There is no immediately apparent Leibniz product rule, or chain rule. In particular, the usual product of two holomorphic functions is not necessarily holomorphic. The Cauchy–Riemann equations imply that f is holomorphic if and only if $\bar{\Delta} f = 0$. If $\bar{\Delta} f = 0$ then $f(z+1) - f(z) = -i(f(z+i) - f(z))$ and, as in classical complex analysis,

$$\Delta f = \Delta_x f = -i \Delta_y f.$$

Definition 2.4. The *interior* of a set $\Omega \subset \mathbb{Z}[i]$ is the subset

$$\mathring{\Omega} = \{z \in \Omega : z+1 \in \Omega,\ z+i \in \Omega\}.$$

Note that for f to be holomorphic at z requires f is defined at z, $z+1$, and $z+i$. Hence, for f to be holomorphic on Ω necessitates that f is defined on G, where $\mathring{G} = \Omega$.

As in the classical case, holomorphic implies infinitely differentiable.

Theorem 2.5. *If f is holomorphic on Ω, then Δf is holomorphic on the interior of Ω.*

Proof. If $z \in \mathring{\Omega}$, then f is holomorphic at z, $z+1$, and $z+i$, so

$$\begin{aligned}
\bar{\Delta} \Delta f(z) &= \bar{\Delta}\left(\frac{f(z+1) - f(z) - i(f(z+i) - f(z))}{2}\right) \\
&= \frac{\bar{\Delta} f(z+1) - \bar{\Delta} f(z) - i(\bar{\Delta} f(z+i) - \bar{\Delta} f(z))}{2} \\
&= 0.
\end{aligned}$$

□

3. Formulas

Theorem 3.1. *Let $z_{n,j} := z + (n-j) + ji$ for $j = 0, \ldots, n$, that is, $\{z_{n,j}\}$ forms the hypotenuse of an isosceles triangle with right angle at z and base length n. If f is holomorphic on the interior of this triangle then*

$$f(z) = \left(\frac{1-i}{2}\right)^n \sum_{j=0}^{n} \binom{n}{j} i^j f(z_{n,j}).$$

Proof. We proceed by induction on n. When $n = 1$, we have

$$f(z) = \frac{1-i}{2}(f(z_{1,0}) + i f(z_{1,1})),$$

which is equivalent to $\bar{\Delta} f(z) = 0$. If f is holomorphic at $z_{n,j}$ then

$$f(z_{n,j}) = \frac{1-i}{2}(f(z_{n+1,j}) + i f(z_{n+1,j+1})).$$

Assuming the formula holds for n we write

$$\begin{aligned}
f(z) &= \left(\frac{1-i}{2}\right)^n \sum_{j=0}^{n} i^j \binom{n}{j} \left(\frac{1-i}{2}\right)(f(z_{n+1,j}) + if(z_{n+1,j+1})) \\
&= \left(\frac{1-i}{2}\right)^{n+1} \left[\sum_{j=0}^{n} i^j \binom{n}{j} f(z_{n+1,j}) + \sum_{j=0}^{n} i^{j+1} \binom{n}{j} f(z_{n+1,j+1}) \right] \\
&= \left(\frac{1-i}{2}\right)^{n+1} \left[\sum_{j=0}^{n} i^j \binom{n}{j} f(z_{n+1,j}) + \sum_{j=1}^{n+1} i^j \binom{n}{j-1} f(z_{n+1,j}) \right] \\
&= \left(\frac{1-i}{2}\right)^{n+1} \left[\sum_{j=0}^{n+1} i^j \binom{n}{j} f(z_{n+1,j}) + \sum_{j=0}^{n+1} i^j \binom{n}{j-1} f(z_{n+1,j}) \right] \\
&= \left(\frac{1-i}{2}\right)^{n+1} \sum_{j=0}^{n+1} i^j f(z_{n+1,j}) \left(\binom{n}{j} + \binom{n}{j-1} \right) \\
&= \left(\frac{1-i}{2}\right)^{n+1} \sum_{j=0}^{n+1} i^j \binom{n+1}{j} f(z_{n+1,j}).
\end{aligned}$$

□

Corollary 3.2. *If* $M_n = \max |f(z + j + ik)|$ *for* $j + k = n$ *and* $j, k \geq 0$,

$$|f(z)| \leq 2^{n/2} M_n.$$

Proof. $|f(z)| \leq (1/\sqrt{2})^n \sum_{j=0}^{n} \binom{n}{j} M_n \leq (1/\sqrt{2})^n 2^n M_n = 2^{n/2} M_n$. □

This formula, unlike the classical Cauchy estimate, grows as $n \to \infty$. So the veracity of Liouville's Theorem in this context remains in doubt. Theorem 3.4 presents a higher-order formula as a consequence of the following lemma.

Lemma 3.3. $$\Delta^k f(z) = \left(\frac{1+i}{2}\right)^k \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(z_{k,j}).$$

Proof. By definition,

$$\Delta^k f(z) = \Delta(\Delta^{k-1} f(z)) = \left(\frac{1+i}{2}\right)(\Delta^{k-1} f(z+1) - \Delta^{k-1} f(z+i)).$$

An induction argument similar to the proof of Theorem 3.1 holds. □

Theorem 3.4. *Let* $z_{n,j} := z + (n - j) + ji$ *for* $j = 0, \ldots, n$. *Then* $\{z_{n,j}\}$ *forms the hypotenuse of an isosceles triangle with right angle at* z *and base length* n. *If* f *is holomorphic on the interior of this triangle then*

$$\Delta^k f(z) = i^k \left(\frac{1-i}{2}\right)^n \sum_{j=0}^{n} \left(i^j f(z_{n,j}) \sum_{l=0}^{k} i^l \binom{k}{l} \binom{n-k}{j-l} \right)$$

for all $n \geq k$.

Proof. Fix k and induct on n. The lemma establishes the case $n = k$. Assuming the formula holds for n we have

$$\begin{aligned}\Delta^k f(z) &= i^k\Big(\frac{1-i}{2}\Big)^{n+1}\sum_{j=0}^{n}\Bigg(i^j\big(f(z_{n+1,j})+if(z_{n+1,j+1})\big)\sum_{l=0}^{k}i^l\binom{k}{l}\binom{n-k}{j-l}\Bigg)\\ &= i^k\Big(\frac{1-i}{2}\Big)^{n+1}\Bigg(\sum_{j=0}^{n+1}i^j f(z_{n+1,j})\sum_{l=0}^{k}i^l\binom{k}{l}\binom{n-k}{j-l}\\ &\qquad\qquad +\sum_{j=0}^{n+1}i^j f(z_{n+1,j})\sum_{l=0}^{k}i^l\binom{k}{l}\binom{n-k}{j-1-l}\Bigg)\\ &= i^k\Big(\frac{1-i}{2}\Big)^{n+1}\sum_{j=0}^{n+1}\Bigg(i^j f(z_{n+1,j})\sum_{l=0}^{k}i^l\binom{k}{l}\binom{n+1-k}{j-l}\Bigg).\end{aligned}$$

□

Theorem 3.1 presents the value of $f(z)$ as a sum of function values along the hypotenuse of the triangle. The following formulas present the value of f at the other triangle vertices as a sum of function values on the opposing side. The proofs are similar to that of Theorem 3.1.

Proposition 3.5. *Let $z_j = z - ni + j$. Then*

$$f(z) = \sum_{j=0}^{n}(1-i)^{n-j}i^j\binom{n}{j}f(z_j).$$

Proposition 3.6. *Let $z_j = z - n + ij$. Then*

$$f(z) = \sum_{j=0}^{n}(1+i)^{n-j}(-i)^j\binom{n}{j}f(z_j).$$

In classical complex analysis, by using Green's theorem, we have a Cauchy formula for continuous, nonholomorphic functions. The discrete analogue of the Cauchy–Pompeiu–Green formula is:

Theorem 3.7. *For any function f defined on the isosceles, right triangle with base length $n \geq 1$,*

$$f(z) = \Big(\frac{1-i}{2}\Big)^n\Bigg(\sum_{j=0}^{n}i^j\binom{n}{j}f(z_{n,j}) - (1-i)\sum_{l=0}^{n-1}\sum_{k=0}^{l}i^k(1+i)^{n-l}\binom{l}{k}\bar{\Delta}f(z_{l,k})\Bigg),$$

where $z_{n,j} = z + (n-j) + ij$.

Proof. We proceed by induction on n. For $n = 1$, we need

$$f(z) = \frac{1-i}{2}\big(f(z_{1,0}) + if(z_{1,1}) - (1-i)(1+i)\bar{\Delta}f(z_{0,0})\big),$$

which holds by the definition of $\bar{\Delta}$. In general, from our base case,

$$f(z_{n,j}) = \frac{1-i}{2}\big(f(z_{n+1,j}) + if(z_{n+1,j+1}) - 2\bar{\Delta}f(z_{n,j})\big).$$

the induction hypothesis,

$$\begin{aligned} f(z) &= \left(\frac{1-i}{2}\right)^n \left(\sum_{j=0}^{n} i^j \binom{n}{j} \left(\frac{1-i}{2}\right) \big(f(z_{n+1,j}) + i f(z_{n+1,j+1}) - 2\bar{\Delta} f(z_{n,j}) \big) \right. \\ &\qquad \left. - (1-i) \sum_{l=0}^{n-1} \sum_{k=0}^{l} i^k (1+i)^{n-l} \binom{l}{k} \bar{\Delta} f(z_{l,k}) \right) \\ &= \left(\frac{1-i}{2}\right)^{n+1} \left(\sum_{j=0}^{n+1} i^j \binom{n+1}{j} f(z_{n+1,j}) - 2 \sum_{j=0}^{n} i^j \binom{n}{j} \bar{\Delta} f(z_{n,j}) \right. \\ &\qquad \left. - 2 \sum_{l=0}^{n-1} \sum_{k=0}^{l} i^k (1+i)^{n-l} \binom{l}{k} \bar{\Delta} f(z_{l,k}) \right) \\ &= \left(\frac{1-i}{2}\right)^{n+1} \left(\sum_{j=0}^{n+1} i^j \binom{n+1}{j} f(z_{n+1,j}) \right. \\ &\qquad \left. - (1-i) \sum_{l=0}^{n} \sum_{k=0}^{l} i^k (1+i)^{n+1-l} \binom{l}{k} \bar{\Delta} f(z_{l,k}) \right). \quad \square \end{aligned}$$

4. Discretization of polynomials

As in the study of discrete calculus of a real variable, we redefine powers so that the power rule holds. In the real variable case, if we consider falling powers $x^{\underline{0}} = 1$ and $x^{\underline{n}} = x(x-1)(x-2)\cdots(x-n+1)$ for $n \geq 1$, the discrete derivative power rule follows:

Proposition 4.1. $\Delta_x x^{\underline{n}} = n x^{\underline{n-1}}$.

Proof.
$$\begin{aligned} \Delta_x x^{\underline{n}} &= (x+1)^{\underline{n}} - x^{\underline{n}} \\ &= (x+1)x \cdots (x-n+2) - x(x-1)\cdots(x-n+1) \\ &= (x+1-(x-n+1))x(x-1)\cdots(x-n+2) = n x^{\underline{n-1}}. \end{aligned}$$ $\square$

To discretize z^n in the complex setting, first expand $z^n = (x+iy)^n$ in terms of x and y and replace each x^n with $x^{\underline{n}}$ and each y^n with $y^{\underline{n}}$. We will denote this polynomial as $z^{\underline{n}}$ or $\mathcal{D}(z^n)$. Hence, our formal definition is

$$\mathcal{D}(z^n) = z^{\underline{n}} = \sum_{k=0}^{n} \binom{n}{k} x^{\underline{n-k}} y^{\underline{k}} i^k.$$

Similarly, the discretization of a polynomial $p(z)$ will be denoted $\mathcal{D}(p(z))$. These *complex falling powers* of z satisfy both the Cauchy–Riemann equations and the following power rule.

Theorem 4.2. $\Delta(z^{\underline{n}}) = n(z^{\underline{n-1}})$ *and* $\bar{\Delta}(z^{\underline{n}}) = 0$.

Proof. Considering the binomial expansion of $z^{\underline{n}}$, by Proposition 4.1,

$$\Delta_x z^{\underline{n}} = \sum_{k=0}^{n} \binom{n}{k} (n-k) x^{\underline{n-k-1}} y^{\underline{k}} i^k \quad \text{and} \quad \Delta_y z^{\underline{n}} = \sum_{k=0}^{n} \binom{n}{k} x^{\underline{n-k}} k y^{\underline{k-1}} i^k$$

and by a change of indices,

$$\Delta_y z^{\underline{n}} = \sum_{k=0}^{n-1} \binom{n}{k+1} x^{\underline{n-k-1}} (k+1) y^{\underline{k}} i^{k+1}.$$

We can simplify these expressions because

$$(n-k)\binom{n}{k} = \frac{(n-k)n!}{k!(n-k)!} = \frac{n(n-1)!}{k!(n-k-1)!} = n\binom{n-1}{k}$$

and

$$(k+1)\binom{n}{k+1} = \frac{(k+1)n(n-1)!}{(n-k-1)!(k+1)k!} = n\binom{n-1}{k}.$$

Using the definition for Δ and simplifying gives

$$\Delta z^{\underline{n}} = \frac{\sum_{k=0}^{n-1} 2n\binom{n-1}{k} x^{\underline{n-k-1}} y^{\underline{k}} i^k}{2} = n z^{\underline{n-1}}.$$

Similarly the definition of $\bar{\Delta}$ gives $\bar{\Delta} z^{\underline{n}} = 0$. □

Corollary 4.3. *If $p(z)$ is a polynomial then $\mathcal{D}(p'(z)) = \Delta(\mathcal{D}(p(z)))$.*

Proof. In both cases the derivative operators are linear. □

In the real case, $x^{\underline{n}} = \sum_{j=0}^{n} s(n, j) x^j$ where $s(n, j)$ are Stirling numbers of the first kind, so we also have the formula

$$z^{\underline{n}} = \sum_{k=0}^{n} i^k \binom{n}{k} \Big(\sum_{j=0}^{n-k} s(n-k, j) x^j \Big) \Big(\sum_{l=0}^{k} s(k, l) y^l \Big).$$

Note that if $n > 1$ then $z^{\underline{n}}$ is not holomorphic in the classical sense.

The definition of complex falling powers may seem unmotivated, so we furnish an example. Consider the difference equation $\Delta F(z) = 2z$. In accordance with the power rule, the solution should be an analogue of z^2. The function F must be of the form

$$z^2 + A\bar{z}^2 - \frac{(1+i)(1+A)}{2} z + B\bar{z} + C,$$

and so

$$\bar{\Delta} F(z) = 2A\bar{z} + \frac{(1+A)(1-i)}{2} + B.$$

Examples of solutions include:

$$\frac{z(z-1) + z(z-i)}{2} + C \quad \text{and} \quad z^2 - \frac{1+i}{2} z - \frac{1-i}{2} \bar{z} + C$$

with the latter being the general solution with $\bar{\Delta} F = 0$; the particular holomorphic solution with $C = 0$ is what we've denoted $z^{\underline{2}}$.

Proposition 4.4. *$\{a + bi : a, b \geq 0$ and $a + b < n\}$ are zeros of $z^{\underline{n}}$.*

Proof. If for each $k = 0, \ldots, n$ we have either $x^{\underline{n-k}} = 0$ or $y^{\underline{k}} = 0$, then

$$z^{\underline{n}} = \sum_{k=0}^{n} i^k \binom{n}{k} x^{\underline{n-k}} y^{\underline{k}} = 0.$$

The zeros of $x^{\underline{j}}$ are given precisely by $\{x \in \mathbb{Z} : 0 \leq x < j\}$ since

$$x^{\underline{j}} = x(x-1)\cdots(x-j-1).$$

So $z^{\underline{n}} = 0$ if for each $k = 0, \ldots, n$ we have either $0 \leq x < n-k$ or $0 \leq y < k$. This condition is met precisely if $x \geq 0,\ y \geq 0$, and $x + y < n$. □

5. Power series and continuation

Lemma 5.1 (Weak Identity Theorem). *If f and g are holomorphic functions which, for some z_0 agree on the line* $\operatorname{Im} z = \operatorname{Im} z_0$, $\operatorname{Re} z \geq \operatorname{Re} z_0$ *then f and g agree for all z such that* $\operatorname{Re} z \geq \operatorname{Re} z_0$ *and* $\operatorname{Im} z \geq \operatorname{Im} z_0$.

Proof. Without loss of generality, assume $z_0 = 0$. If f and g agree on the positive real line, then, since both are holomorphic, they have a unique holomorphic extension to the points above this line. □

The standard Schwarz reflection principle for holomorphic continuation does not hold. Falling power series can be represented as

$$\sum_{n=0}^{\infty} a_n (z - z_0)^{\underline{n}}.$$

Regions of convergence have vastly different shapes from those in the classical case.

Theorem 5.2. *The falling power series*

$$\sum_{n=0}^{\infty} a_n (z - z_0)^{\underline{n}}$$

converges for at least all z such that $\operatorname{Re} z \geq \operatorname{Re} z_0$ *and* $\operatorname{Im} z \geq \operatorname{Im} z_0$.

Proof. We will prove this for $z_0 = 0$ and the proof can be carried out similarly for other finite z_0. By Proposition 4.4 the zeros of $z^{\underline{n}}$ include

$$\{a + bi : a, b \geq 0 \text{ and } a + b < n\}.$$

For any point $a + bi$ in the first quadrant, there exists n with $a + b < n$. Thus the terms of the series $z^{\underline{k}}$ with $k > n$ will be 0 for $z = a + bi$. A sum of a finite number of terms is trivially convergent. Since $a + bi$ was arbitrary, the falling power series converges for every point in the first quadrant. □

The first quadrant may not be the only place a falling power series centered at 0 converges. For instance, the series

$$\sum_{n=0}^{\infty} \frac{z^{\underline{n}}}{(n+1)!}$$

evaluated at $z = -1$ is the alternating harmonic series and thus converges to $\ln 2$.

Proposition 5.3. *If the falling power series*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^{\underline{n}}$$

converges on a domain Ω, *then*

$$\Delta f(z) = \sum_{n=0}^{\infty} \Delta\left(a_n (z - z_0)^{\underline{n}}\right) = \sum_{n=1}^{\infty} n a_n (z - z_0)^{\underline{n-1}}$$

for $z \in \mathring{\Omega}$.

Proof. For any point $z \in \mathring{\Omega}$, the series converges at $\{z, z+1, z+i\}$. So

$$\begin{aligned}
\Delta f(z) &= \frac{f(z+1) - f(z) - i\left(f(z+i) - f(z)\right)}{2} \\
&= \frac{\sum a_n (z+1)^{\underline{n}} - \sum a_n (z)^{\underline{n}} - i\left(\sum a_n (z+i)^{\underline{n}} - \sum a_n (z)^{\underline{n}}\right)}{2} \\
&= \sum a_n \left(\frac{(z+1)^{\underline{n}} - z^{\underline{n}} - i\left((z+i)^{\underline{n}} - z^{\underline{n}}\right)}{2} \right) \\
&= \sum a_n \Delta z^{\underline{n}}.
\end{aligned}$$

□

Definition 5.4. A function is *analytic* if it can be written locally as a convergent falling power series.

Proposition 5.5. *Analytic on* Ω *implies holomorphic on* $\mathring{\Omega}$.

Proof. As in the proof of Proposition 5.3, for $z \in \mathring{\Omega}$, the $\bar{\Delta}$ can be applied to the series term by term. For each n, $\bar{\Delta} z^{\underline{n}} = 0$ and so $\bar{\Delta} f(z) = 0$. Thus f is holomorphic on $\mathring{\Omega}$. □

This brings us one of the main results dealing with falling power series.

Theorem 5.6. *Holomorphic implies analytic.*

Proof. We may assume $z_0 = 0$ and the series converges everywhere in the first quadrant (Theorem 5.2). By interpolation, we can form a unique falling power

series which agrees with the function f on the positive real line according to the recurrence relations

$$a_0 = f(0) \quad \text{and} \quad a_n = \frac{f(n) - \sum_{k=0}^{n-1} k^{\underline{n}} a_k}{n^{\underline{n}}}.$$

From Proposition 5.5, we know that the series is holomorphic, and by the weak identity theorem, since f agrees with this power series on the real line, then it agrees with the series in the whole first quadrant, and f is analytic there. □

From the proof of Theorem 5.6, follows the usual Taylor expression.

Corollary 5.7 (Taylor's theorem). *A holomorphic function f is locally given by the falling power series*

$$f(z) = \sum_{n=0}^{\infty} \frac{\Delta^n f(z_0)}{n!} (z - z_0)^{\underline{n}}.$$

6. Elementary functions

First, a discrete analogue of the exponential function:

Proposition 6.1. *If $\Delta f = f$ and $\bar{\Delta} f = 0$, then*

$$f(x + iy) = 2^x (1 + i)^y f(0).$$

Proof. Setting $f(z) = \Delta f(z)$ gives

$$f(z) = \frac{f(z+1) - f(z) - i(f(z+i) - f(z))}{2},$$

and $\bar{\Delta} f(z) = 0$ gives

$$\frac{f(z+1) - f(z) + i(f(z+i) - f(z))}{2} = 0.$$

After some simplification, we obtain

$$f(z+1) = 2f(z) \quad \text{and} \quad f(z+i) = (1+i) f(z).$$

With these two functional equations,

$$f(x + iy) = 2^x f(iy) = 2^x (1 + i)^y f(0).$$ □

Definition 6.2. The *discrete complex exponential* is given by

$$\exp(z) = \exp(x + iy) = 2^x (1 + i)^y.$$

Note that it satisfies a law of exponents, i.e., $\exp(z+w) = \exp(z)\exp(w)$. As a falling power series, for z in the first quadrant,

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^{\underline{n}}}{n!}.$$

Analogous to classical complex analysis where $e^{2\pi ik} = 1$, we have:

Proposition 6.3. $\exp(z) = 1$ *if and only if* $z = (4-8i)k$ *for some integer* k.

Proof. $\arg(2^x(1+i)^y) = y \cdot \arg(1+i) = y\pi/4$, which is a multiple of 2π if and only if y is a multiple of 8. Next, $|\exp(x+iy)| = |2^x| \cdot |1+i|^y = 2^{x+y/2}$, which equals 1 if and only if $2x = -y$. We may conclude that $\exp(x+iy) = 1$ if and only if $x+iy = (4-8i)k$ for some integer k. □

Next, we look for analogues of sine and cosine.

Proposition 6.4. *If* $-\Delta^2 f = f$ *and* $\bar{\Delta} f = 0$ *then*

$$f(x+iy) = (1-i)^x 2^y f(0).$$

Proof. If f is holomorphic at z, then $\Delta f(z) = \Delta_x f(z) = -i\,\Delta_y f(z)$. Hence,

$$-\Delta^2 f(z) = -\Delta_y \Delta_y f(z) = \Delta_y f(z+i) - \Delta_y f(z) = f(z+2i) - 2f(z+i) + f(z).$$

Setting equal to $f(z)$ gives $f(z+2i) = 2f(z+i)$, or by change of variables

$$f(z+i) = 2f(z).$$

Also,

$$\begin{aligned} -\Delta^2 f(z) = i\,\Delta_y \Delta_x f(z) &= i\,\Delta_y f(z+1) - i\,\Delta_y f(z) \\ &= if(z+1+i) - if(z+1) - if(z+i) + if(z). \end{aligned}$$

Setting equal to $f(z)$ and substituting $f(z+1+i) = 2f(z+1)$ and $f(z+i) = 2f(z)$ gives $(1+i)f(z) = if(z+1)$, or

$$f(z+1) = (1-i)f(z).$$

Combining results yields the solution $f(x+iy) = (1-i)^x 2^y f(0)$. □

Motivated by the classical equation, $e^{x+iy} = e^x(\cos y + i \sin y)$, let us find an analogue for $\exp(x+iy) = 2^x(1+i)^y$, by setting

$$c(t) = \mathrm{Re}(1+i)^t \quad \text{and} \quad s(t) = \mathrm{Im}(1+i)^t$$

for $t \in \mathbb{Z}$. With these definitions on the real line, define $c(x+iy)$ and $s(x+iy)$ for $y \geq 0$ by holomorphic extension to the upper half-plane.

Proposition 6.5. *For $y > 0$,*

$$c(x+iy) = \frac{(1-i)^x 2^y}{2} \quad \textit{and} \quad s(x+iy) = \frac{(1-i)^x 2^y}{-2i}.$$

Proof. By Proposition 6.4, the functions

$$(x+iy) \mapsto \frac{(1-i)^x 2^y}{2} \quad \text{and} \quad (x+iy) \mapsto \frac{(1-i)^x 2^y}{-2i}$$

are holomorphic everywhere. Hence by Lemma 5.1, it is sufficient to show that equality holds on the line $\operatorname{Im} z = 1$. Let $x \in \mathbb{Z}$. Since c is holomorphic at x, by Proposition 3.5,

$$\begin{aligned} c(x+i) &= (1-i)c(x) + ic(x+1) \\ &= (1-i)\operatorname{Re}(1+i)^x + i\operatorname{Re}(1+i)^{x+1} \\ &= (1-i)\operatorname{Re}(1+i)^x + i(\operatorname{Re}(1+i)^x - \operatorname{Im}(1+i)^x) \\ &= \operatorname{Re}(1+i)^x - i\operatorname{Im}(1+i)^x = \overline{(1+i)^x} = (1-i)^x, \end{aligned}$$

which equals $\frac{(1-i)^x 2^y}{2}$ for $y = 1$. Similarly,

$$\begin{aligned} s(x+i) &= (1-i)s(x) + is(x+1) \\ &= (1-i)\operatorname{Im}(1+i)^x + i\operatorname{Im}(1+i)^{x+1} \\ &= (1-i)\operatorname{Im}(1+i)^x + i(\operatorname{Re}(1+i)^x + \operatorname{Im}(1+i)^x) \\ &= i\operatorname{Re}(1+i)^x + \operatorname{Im}(1+i)^x = i\overline{(1+i)^x} = i(1-i)^x, \end{aligned}$$

which equals $\frac{(1-i)^x 2^y}{-2i}$ for $y = 1$. □

7. Path integration

Definition 7.1. A *path* γ of length n is a sequence $\{\gamma_j\}_{j=0}^n \subset \mathbb{Z}[i]$ such that

$$|\gamma_j - \gamma_{j-1}| = 1,$$

for every integer j such that $1 \le j \le n$. A closed path satisfies $\gamma_0 = \gamma_n$.

Definition 7.2. A *simply connected domain* Ω is a path-connected set of points $\{z \in \mathbb{Z}[i]\}$ with no holes, i.e., Ω is such that the interior of every closed path set lies inside Ω.

Definition 7.3. A *corner* of a path γ is a point γ_j with $0 < j < n$ such that

$$|\gamma_{j+1} - \gamma_{j-1}| \neq 2.$$

Definition 7.4. The *path integral* of f along γ is

$$\int_\gamma f(z) = \sum_{j=1}^{n} f\big(\min\{x_j, x_{j-1}\} + i\min\{y_j, y_{j-1}\}\big)(\gamma_j - \gamma_{j-1}),$$

where $x_j = \operatorname{Re}\gamma_j$ and $y_j = \operatorname{Im}\gamma_j$ for $0 \le j \le n$.

Lemma 7.5. *If γ is a path from γ_0 to γ_n with no corners and f is holomorphic everywhere along the path, then*

$$\int_\gamma \Delta f(z) = f(\gamma_n) - f(\gamma_0).$$

Proof. For a horizontal path oriented from left to right having no corners, $\gamma_j - \gamma_{j-1}$ is constant and equal to 1, so

$$\int_\gamma \Delta f(z) = \int_\gamma \Delta_x f(z) = \int_\gamma f(z+1) - f(z) = \sum_{j=1}^{n} f(\gamma_j) - f(\gamma_{j-1}),$$

which telescopes leaving $\int_\gamma \Delta f(z) = f(\gamma_n) - f(\gamma_0)$. For a horizontal path oriented from right to left, $\gamma_j - \gamma_{j-1} = -1$, so

$$\int_\gamma \Delta f(z) = -\sum_{j=1}^{n} f(\gamma_{j-1}) - f(\gamma_j) = f(\gamma_n) - f(\gamma_0).$$

For a vertical path oriented from bottom to top, $\gamma_j - \gamma_{j-1} = i$, so

$$\int_\gamma \Delta f(z) = \int_\gamma -i\,\Delta_y f(z) = -i \sum_{j=1}^{n} \big(f(\gamma_j) - f(\gamma_{j-1})\big)\, i = f(\gamma_n) - f(\gamma_0).$$

For a vertical path oriented from top to bottom, $\gamma_j - \gamma_{j-1} = -i$, so

$$\int_\gamma \Delta f(z) = \int_\gamma -i\,\Delta_y f(z) = -i \sum_{j=1}^{n} \big(f(\gamma_{j-1}) - f(\gamma_j)\big)\,(-i) = f(\gamma_n) - f(\gamma_0). \quad \square$$

Theorem 7.6 (Fundamental Theorem). *If γ is a path from γ_0 to γ_n and f is holomorphic everywhere along the path, then*

$$\int_\gamma \Delta f(z) = f(z_n) - f(z_0).$$

Proof. Decompose γ into a union of paths having no corners:

$$\gamma = \gamma^1 + \gamma^2 + \cdots + \gamma^m.$$

Then

$$\begin{aligned}\int_\gamma \Delta f &= \int_{\gamma^1} \Delta f + \int_{\gamma^2} \Delta f + \cdots + \int_{\gamma^m} \Delta f \\ &= (f(\gamma^m_{n_m}) - f(\gamma^m_0)) + \cdots + (f(\gamma^1_{n_1}) - f(\gamma^1_0)) = f(\gamma_n) - f(\gamma_0). \quad \square\end{aligned}$$

Corollary 7.7. *If $\Delta f(z) = 0$ on a path-connected set Ω, then $f(z)$ is constant on Ω.*

Proof. If z and w are in Ω, there exists a path in Ω from z to w. Since $\int_\gamma \Delta f(z) = 0$, it follows that $f(w) = f(z)$. □

Lemma 7.8 (Goursat's lemma). *Let T be a unit square given by the path*

$$\{z,\ z+1,\ z+1+i,\ z+i,\ z\},$$

and suppose f is holomorphic at z. Then $\int_T f(z) = 0$.

Proof. $\int_T f(z) = f(z) + if(z+1) - f(z+i) - if(z) = 2i\,\bar{\Delta} f(z) = 0.$ □

The following corollary immediately follows.

Corollary 7.9 (Morera's theorem). *Let f be a function defined on a set G. If $\int_T f(z) = 0$ for all unit squares T whose interior point is contained in the interior of G, then f is holomorphic on the interior of G.*

Theorem 7.10 (Cauchy's theorem). *Let Ω be a simply connected domain and let γ be a closed path in Ω. Then*

$$\int_\gamma f(z) = 0,$$

for each function f that is holomorphic on Ω.

Proof. γ can be written as a canceling sum of unit squares, $T_1 + T_2 + \cdots + T_m$. Since Ω is simply connected, all of these squares lie in the interior of Ω. By Goursat's lemma, $\int_{T_1} f(z) = \int_{T_2} f(z) = \cdots = \int_{T_n} f(z) = 0$, and so

$$\int_\gamma f(z) = \int_{T_1} f(z) + \int_{T_2} f(z) + \cdots + \int_{T_n} = 0.$$ □

Theorem 7.11. *If f is holomorphic on a simply connected domain, Ω, then f has a primitive in Ω.*

Proof. Fix $z_0 \in \Omega$. Let $F(z) = \int_\gamma f(w)$ where $\gamma_0 = z_0$ and $\gamma_n = z$. By Cauchy's Theorem, this function is path-independent and well-defined.

$$\begin{aligned}
\Delta F(z) &= \Delta \int_\gamma f(w) \\
&= \frac{\int_{z_0}^{z+1} f(w) - \int_{z_0}^{z} f(w) - i\left(\int_{z_0}^{z+i} f(w) - \int_{z_0}^{z} f(w)\right)}{2} \\
&= \frac{\int_{z_0}^{z} f(w) + f(z) - \int_{z_0}^{z} f(w) - i\left(\int_{z_0}^{z} f(w) + if(z) - \int_{z_0}^{z} f(w)\right)}{2} \\
&= f(z).
\end{aligned}$$ □

Acknowledgments

Funding for this undergraduate research was provided by the Taylor University SRTP Mini Grant. Only after these results were formulated did the authors find the pertinent literature on monodiffric functions.

References

[Bohner and Peterson 2001] M. Bohner and A. Peterson, *Dynamic equations on time scales*, Birkhäuser, Boston, MA, 2001. MR 2002c:34002 Zbl 0978.39001

[Duffin 1956] R. J. Duffin, "Basic properties of discrete analytic functions", *Duke Math. J.* **23** (1956), 335–363. MR 17,1193e Zbl 0070.30503

[Ferrand 1944] J. Ferrand, "Fonctions préharmoniques et fonctions préholomorphes", *Bull. Sci. Math.* (2) **68** (1944), 152–180. MR 7,149g Zbl 0063.01349

[Isaacs 1941] R. P. Isaacs, "A finite difference function theory", *Univ. Nac. Tucumán Rev. A* **2** (1941), 177–201. MR 3,298d Zbl 0061.15902

[Isaacs 1952] R. Isaacs, "Monodiffric functions", pp. 257–266 in *Construction and applications of conformal maps*, National Bureau of Standards Appl. Math. Ser. **18**, US Government Printing Office, Washington, DC, 1952. MR 14,633g

[Kiselman 2005] C. O. Kiselman, "Functions on discrete sets holomorphic in the sense of Isaacs, or monodiffric functions of the first kind", *Sci. China Ser. A* **48**:suppl. (2005), 86–96. MR 2006d:30069

[Mercat 2001] C. Mercat, "Discrete Riemann surfaces and the Ising model", *Comm. Math. Phys.* **218**:1 (2001), 177–216. MR 2002c:82019 Zbl 1043.82005

Received: 2009-09-24 Revised: 2010-09-22 Accepted: 2010-09-23

jseaborn@email.unc.edu *Mathematics Department, The University of North Carolina, CB #3250, Phillips Hall 412, Chapel Hill, NC 27599, United States*

phmummert@taylor.edu *Mathematics Department, Taylor University, 236 W Reade Ave, Upland, IN 46989, United States*
http://faculty.taylor.edu/phmummert/

INVOLVE 3:3(2010)

$\zeta(n)$ via hyperbolic functions

Joseph D'Avanzo and Nikolai A. Krylov

(Communicated by Ken Ono)

We present an approach to compute $\zeta(2)$ by changing variables in the double integral using hyperbolic trigonometric functions. We also apply this approach to present $\zeta(n)$, when $n > 2$, as a definite improper integral of a single variable.

1. Introduction

The Riemann zeta function is defined as the series

$$\zeta(n) = \frac{1}{1^n} + \frac{1}{2^n} + \frac{1}{3^n} + \cdots + \frac{1}{k^n} + \cdots$$

for any integer $n \geq 2$. Three centuries ago Euler found that $\zeta(2) = \pi^2/6$, which is an irrational number. The exact value of $\zeta(3)$ is still unknown, though it was proved by Apéry in 1979 that $\zeta(3)$ is also irrational [van der Poorten 1979]. The values of $\zeta(n)$, when n is even, are known and can be written in terms of Bernoulli numbers. We refer the interested reader to Chapter 19 of [Aigner and Ziegler 2001] for a "perfect" proof of the formula

$$\zeta(2k) = \sum_{n=1}^{\infty} \frac{1}{n^{2k}} = \frac{(-1)^{k-1} 2^{2k-1} B_{2k}}{(2k)!} \cdot \pi^{2k} \quad (k \in \mathbb{N}).$$

Notice that $\zeta(n)$ can be written as the following multivariable integral:

$$\zeta(n) = \int_0^1 \cdots \int_0^1 \frac{1}{1 - x_1 x_2 \cdots x_n} \, dx_1 \, dx_2 \ldots \, dx_n.$$

Indeed, each integral is improper at both ends and since the geometric series

$$\sum_{q \geq 0} x^q$$

MSC2000: primary 26B15; secondary 11M06.
Keywords: multiple integrals, Riemann's zeta function.

converges uniformly on the interval $|x| \le R$ for all $R \in (0, 1)$, we can write

$$\frac{1}{1 - x_1 x_2 \cdots x_n} = \sum_{q=0}^{\infty} (x_1 x_2 \cdots x_n)^q,$$

then interchange summation with integration, and then integrate $(x_1 x_2 \cdots x_n)^q$ for each q. Using the identities

$$\frac{1}{1-xy} + \frac{1}{1+xy} = \frac{2}{1-x^2y^2} \quad \text{and} \quad \frac{1}{1-xy} - \frac{1}{1+xy} = \frac{2xy}{1-x^2y^2}$$

and a simple change of variables, one can easily see that

$$\int_0^1 \int_0^1 \frac{1}{1-xy}\, dx\, dy = \frac{4}{3} \int_0^1 \int_0^1 \frac{1}{1-x^2y^2}\, dx\, dy.$$

By further generalizing this idea, one reaches

$$\zeta(n) = \frac{2^n}{2^n - 1} \int_0^1 \cdots \int_0^1 \frac{1}{1 - \prod_{i=1}^n x_i^2}\, dx_1 \ldots dx_n.$$

Notice that $(1, 1)$ is the only point in the square $[0, 1] \times [0, 1]$ that makes the integrand $1/(1-x^2y^2)$ singular. If we take another point on the graph of $1 = x^2y^2$, say, $(a, 1/a)$ with $a \in (0, \infty)$, it follows easily (see Lemma 1.1 below) that

$$\int_0^{1/a} \int_0^a \frac{1}{1-x^2y^2}\, dx\, dy = \int_0^1 \int_0^1 \frac{1}{1-x^2y^2}\, dx\, dy.$$

This result motivates the following definition:

Definition 1. For any point $(a_1, a_2, \ldots, a_{n-1}) \in \mathbb{R}^{n-1}$ such that $a_i \in (0, +\infty)$, for all $i \in \{1, \ldots, n-1\}$ we define

$$I_n(a_1, \ldots, a_{n-1}) = \int_0^{1/(a_1 \cdots a_{n-1})} \cdots \int_0^{a_2} \int_0^{a_1} \frac{1}{1 - \prod_{i=1}^n x_i^2}\, dx_1\, dx_2 \ldots dx_n.$$

Lemma 1.1. *For any $a_i \in (0, +\infty)$, we have $I_n(a_1, \ldots, a_{n-1}) = I_n(1, 1, \ldots, 1)$.*

Proof. Simply observe that by using the change of variables

$$x_i = a_i u_i \quad \text{for all } i \in \{1, \ldots, n\}, \text{ where } a_n = 1/(a_1 a_2 \cdots a_{n-1}),$$

the Jacobian equals 1, and the integrand is unchanged. □

In this article we investigate $\zeta(n)$ following Beukers, Calabi and Kolk [1993], who used the change of variables

$$x = \frac{\sin u}{\cos v} \quad \text{and} \quad y = \frac{\sin v}{\cos u} \quad \text{to evaluate} \int_0^1 \int_0^1 \frac{1}{1-x^2y^2}\, dx\, dy.$$

A similar proof of the identity $\zeta(2) = \pi^2/6$ may also be found in [Aigner and Ziegler 2001, Chapter 6] and [Elkies 2003; Kalman 1993]. This last reference in addition to a few other proofs of the identity, contains a history of the problem and an extensive reference list.

Here we will be changing variables too, but in the integrals $I_n(a_1, \ldots, a_{n-1})$ and using the hyperbolic trigonometric functions sinh and cosh instead of sin and cos. Such a change of variables was considered independently by [Silagadze 2010].

2. Hyperbolic change of variables

The change of variables

$$x_i = \frac{\sin u_i}{\cos u_{i+1}}, \qquad \text{for all } i \in \mathbb{N} \bmod n$$

reduces the integrand in $I_n(1, \ldots, 1)$ to 1 only when n is even. The region of integration $\Phi_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : 0 < x_1, \ldots, x_n < 1\}$ becomes the one-to-one image of the n-dimensional polytope

$$\Pi_n := \left\{(u_1, u_2, \ldots, u_n) \in \mathbb{R}^n : u_i > 0,\ u_i + u_{i+1} < \frac{\pi}{2},\ 1 \le i \le n\right\}$$

(note that $u_{n+1} = u_1$).

We suggest here a different change of variables that will produce an integrand of 1 for all values of n in $I_n(a_1, \ldots, a_{n-1})$. But first we define the corresponding region.

Definition 2. For any point $(a_1, a_2, \ldots, a_{n-1}) \in \mathbb{R}^{n-1}$ such that $a_i \in (0, +\infty)$, for all $i \in \{1, \ldots, n-1\}$, we define

$$\Phi_n(a_1, a_2, \ldots, a_{n-1}) := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid 0 < x_i < a_i \text{ for all } i \in \{1, \ldots, n\}\},$$

where $a_n = 1/(a_1 \cdot a_2 \cdot \ldots \cdot a_{n-1})$.

Lemma 2.1. *The change in variables*

$$x_i = \frac{\sinh u_i}{\cosh u_{i+1}} \qquad \textit{for all } i \in \mathbb{N} \bmod n$$

reduces the integrand of $I_n(a_1, \ldots, a_{n-1})$ to 1 for all values of $n \ge 2$. It also gives a one-to-one differentiable map between the region $\Phi_n(a_1, a_2, \ldots, a_{n-1})$ and the set $\Gamma_n \subset \mathbb{R}^n$ described by the n inequalities

$$0 < u_i < \operatorname{arcsinh}(a_i \cosh u_{i+1}), \qquad \textit{for all } i \in \mathbb{N} \bmod n.$$

The set Γ_2 is illustrated in Figure 1.

Proof. The inequalities for Γ_n follow trivially from the corresponding inequalities $0 < x_i < a_i$ and the facts that $\cosh x > 0$ and $\operatorname{arcsinh} x$ is increasing everywhere. Injectivity and smoothness of the map may be proven by writing down formulas, which express each u_i in terms of all x_j. For example, here are the corresponding formulas for the set Γ_3:

$$u_i = \operatorname{arcsinh}\left(x_i \sqrt{\frac{1 + x_{i+1}^2 + x_{i-1}^2 x_{i+1}^2}{1 - x_1^2 x_2^2 x_3^2}} \right), \quad i \in \mathbb{N} \bmod 3.$$

The Jacobian is the determinant of the matrix

$$A = \begin{pmatrix} \dfrac{\cosh u_1}{\cosh u_2} & \dfrac{-\sinh u_1 \sinh u_2}{\cosh^2 u_2} & 0 & \dots & 0 \\ 0 & \dfrac{\cosh u_2}{\cosh u_3} & \dfrac{-\sinh u_2 \sinh u_3}{\cosh^2 u_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{-\sinh u_n \sinh u_1}{\cosh^2 u_1} & 0 & 0 & \dots & \dfrac{\cosh u_n}{\cosh u_1} \end{pmatrix}.$$

To compute this determinant we observe that the first column expansion reduces the computation to two determinants of the upper and lower triangular matrices. This results in the formula below, where the first term comes from the upper triangular matrix and the second from the lower triangular matrix (recall that $u_{n+1} = u_1$):

$$\det A = \prod_{i=1}^{n} \frac{\cosh u_i}{\cosh u_{i+1}} + (-1)^{n-1} \cdot \prod_{i=1}^{n} \frac{-\sinh u_i \sinh u_{i+1}}{\cosh^2 u_{i+1}} = 1 - \prod_{i=1}^{n} \tanh^2 u_i.$$

When using the above change in variables, the denominator of the integrand $1 - \prod_{i=1}^{n} x_i^2$ becomes $1 - \prod_{i=1}^{n} \tanh^2 u_i$, which we just proved to be the Jacobian. □

3. Computation of $\zeta(2)$

We begin with $\zeta(2)$, which is a rational multiple of $I_2(1)$. Lemma 1.1 implies that it's enough to compute

$$I_2(a) = \int_0^{1/a} \int_0^{a} \frac{1}{1 - x^2 y^2}\, dx\, dy \quad \text{for arbitrary } a > 0.$$

We now perform the change in variables

$$x = \frac{\sinh u}{\cosh v}, \quad y = \frac{\sinh v}{\cosh u}.$$

As we proved above, our integrand reduces to 1 and all we must do is worry about the limits. If $x = 0$, then clearly $u = 0$; the same is true for y and v. If $x = a$ then $a \cosh v = \sinh u$, so $v = \operatorname{arccosh}(\sinh(u)/a)$, and if $y = 1/a$, then

$$\frac{1}{a} \cosh u = \sinh v,$$

so $v = \operatorname{arcsinh}(\cosh(u)/a)$ — thus describing our region of integration, depicted in Figure 1. We then write the integral $I_2(a)$ as

$$\int_0^{\operatorname{arcsinh} a} \operatorname{arcsinh} \frac{\cosh u}{a}\, du + \int_{\operatorname{arcsinh} a}^{\infty} \left(\operatorname{arcsinh} \frac{\cosh u}{a} - \operatorname{arccosh} \frac{\sinh u}{a} \right) du.$$

Lemma 3.1. $$\lim_{a \to 0} \int_0^{\operatorname{arcsinh} a} \operatorname{arcsinh} \frac{\cosh u}{a}\, du = 0.$$

Proof. We have $\cosh \operatorname{arcsinh} z = \sqrt{1 + z^2}$. Therefore

$$\operatorname{arcsinh} \frac{\cosh \operatorname{arcsinh} a}{a} = \operatorname{arcsinh} \sqrt{\frac{1}{a^2} + 1}\,.$$

Since $\operatorname{arcsinh}(\cosh(u)/a)$ is convex, we can take the area of the rectangle with vertices at $(0, 0)$, $(\operatorname{arcsinh} a, 0)$, and $(\operatorname{arcsinh} a, \operatorname{arcsinh}(\cosh(\operatorname{arcsinh}(a))/a)$ as an overestimate of the integral — that is,

$$\operatorname{arcsinh} a \ \operatorname{arcsinh} \sqrt{\frac{1}{a^2} + 1} \geq \int_0^{\operatorname{arcsinh} a} \operatorname{arcsinh} \frac{\cosh u}{a}\, du \geq 0.$$

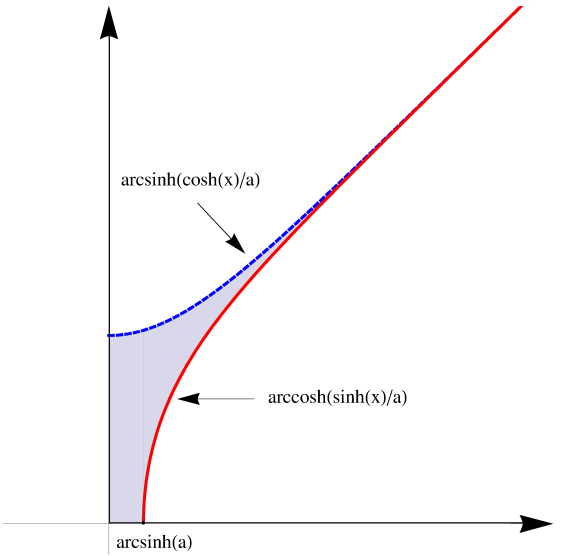


Figure 1. The set $\Gamma_2 \subset \mathbb{R}^2$, for all $a > 0$.

Then by applying L'Hospital's rule one can deduce

$$\lim_{a\to 0} \operatorname{arcsinh} a \cdot \operatorname{arcsinh} \sqrt{\frac{1}{a^2}+1} = 0. \qquad \square$$

Now, since $I_2(a) = I_2(1)$, for all $a > 0$, we conclude that $I_2(1) = \lim_{a\to 0} I_2(a)$. Therefore we have

$$I_2(1) = \lim_{a\to 0} \int_{\operatorname{arcsinh} a}^{\infty} \left(\operatorname{arcsinh} \frac{\cosh u}{a} - \operatorname{arccosh} \frac{\sinh u}{a} \right) du,$$

which, after taking the limit as $a \to 0$, gives

$$I_2(1) = \int_0^{\infty} \ln \frac{\cosh x}{\sinh x}\, dx.$$

Using integration by parts with $u = \ln \dfrac{\cosh x}{\sinh x}$ and $v = dx$ one obtains

$$I_2(1) = x \ln \frac{\cosh x}{\sinh x}\bigg|_0^{\infty} + \int_0^{\infty} \frac{2x}{\sinh 2x}\, dx.$$

By examining the limits of the first half of the formula as x goes to 0 and ∞ we are left with only the integral

$$I_2(1) = \int_0^{\infty} \frac{2x}{\sinh 2x}\, dx.$$

By applying the change in variables $u = 2x$ our formula becomes

$$I_2(1) = \frac{1}{2} \int_0^{\infty} \frac{u}{\sinh u}\, du.$$

Now we use the method of differentiation under the integral sign and consider the function

$$F(\alpha) = \frac{1}{2} \int_0^{\infty} \frac{\operatorname{arctanh}(\alpha \tanh x)}{\sinh x}\, dx.$$

One should consider the function F at the points $\alpha = 1$ and $\alpha = 0$. $F(1)$ is clearly the integral we are trying to find and $F(0)$ is 0. Thus, by differentiating under the integral with respect to α and using some algebra, we obtain

$$F'(\alpha) = f(\alpha) = \frac{1}{2} \int_0^{\infty} \frac{\cosh x}{1 + (1-\alpha^2)\sinh^2 x}\, dx.$$

Then, by performing the change of variables $u = \sqrt{1-\alpha^2} \cdot \sinh x$, the integral becomes

$$f(\alpha) = \frac{1}{2\sqrt{1-\alpha^2}} \int_0^{\infty} \frac{1}{1+u^2}\, du,$$

which is simply

$$\frac{\operatorname{arctanh} u}{2\sqrt{1-\alpha^2}}\bigg|_0^\infty = \frac{\pi}{4\sqrt{1-\alpha^2}}.$$

Since we took the derivative with respect to α we must take the integral with respect to alpha, so we have

$$\int_0^1 f(\alpha)\, d\alpha = F(1) - F(0) = F(1) - 0 = F(1),$$

which, as stated above is our goal. So

$$I_2(1) = \int_0^1 f(\alpha)\, d\alpha = \frac{\pi}{4}\int_0^1 \frac{1}{\sqrt{1-\alpha^2}}\, d\alpha = \frac{\pi}{4}\arcsin\alpha\Big|_0^1 = \frac{\pi^2}{8},$$

and, hence,

$$\zeta(2) = \frac{4}{3}\cdot\frac{\pi^2}{8} = \frac{\pi^2}{6}.$$

4. A formula for $\zeta(n)$, $n \geq 2$

One can try to use a similar approach to compute $\zeta(n)$, for $n > 2$, however the computations become too long. Instead, we present an elementary proof of the following theorem, which generalizes our formula for $\zeta(2)$ from Section 3.

Theorem 4.1. *Let $n \geq 2$ be a natural number. Then*

$$\int_0^1 \cdots \int_0^1 \frac{1}{1-\prod_{i=1}^n x_i^2}\, dx_1 \ldots dx_n = \frac{1}{(n-1)!}\cdot\int_0^\infty \ln^{n-1}(\coth x)\, dx.$$

We start with the following lemma, which can be easily proved by using induction on k, integration by parts and L'Hospital's rule.

Lemma 4.2. $$\int_0^1 \ln^k(z)\, z^{2q} dz = \frac{(-1)^k k!}{(2q+1)^{k+1}} \quad \textit{for all } k \in \mathbb{N} \textit{ and } q \geq 0.$$

Proof of the theorem. Applying the substitution $z = \tanh x$ to the integral

$$\frac{1}{(n-1)!}\int_0^\infty \ln^{n-1}(\coth x)\, dx,$$

gives

$$\frac{1}{(n-1)!}\int_0^1 \frac{(-\ln z)^{n-1}}{1-z^2}\, dz = \frac{1}{(n-1)!}\int_0^1 (-\ln z)^{n-1}\left(\sum_{q\geq 0} z^{2q}\right) dz.$$

Since the integral is improper at both ends and the geometric series $\sum_{q\geq 0} z^{2q}$ converges uniformly on the interval $|z| \leq R$, for all $R \in (0,1)$, the last integral equals

$$\frac{1}{(n-1)!}\sum_{q\geq 0}(-1)^{n-1}\int_0^1 \ln^{n-1}(z) z^{2q} dz,$$

which, by Lemma 4.2, is equal to

$$\sum_{q\geq 0} \frac{1}{(2q+1)^n}.$$

Using the geometric series expansion one can easily show that we also have

$$\int_0^1 \cdots \int_0^1 \frac{1}{1-\prod_{i=1}^n x_i^2}\, dx_1 \ldots\, dx_n = \sum_{q\geq 0} \frac{1}{(2q+1)^n}. \qquad \square$$

Corollary 4.3. *For any integer $n \geq 2$,*

$$\zeta(n) = \frac{2^n}{(2^n-1)(n-1)!} \int_0^\infty \ln^{n-1}(\coth x)\, dx.$$

Acknowledgement

The authors thank the referee, who drew our attention to [Kalman 1993] and made a few useful suggestions that improved the exposition.

References

[Aigner and Ziegler 2001] M. Aigner and G. M. Ziegler, *Proofs from The Book*, 2nd ed., Springer, Berlin, 2001. MR 2001j:00001 Zbl 0978.00002

[Beukers et al. 1993] F. Beukers, J. A. C. Kolk, and E. Calabi, "Sums of generalized harmonic series and volumes", *Nieuw Arch. Wisk.* (4) **11**:3 (1993), 217–224. MR 94j:11022 Zbl 0797.40001

[Elkies 2003] N. D. Elkies, "On the sums $\sum_{k=-\infty}^{\infty}(4k+1)^{-n}$", *Amer. Math. Monthly* **110**:7 (2003), 561–573. MR 2004f:11152 Zbl 1083.11055

[Kalman 1993] D. Kalman, "Six Ways to Sum a Series", *The College Mathematics Journal* **24**:5 (1993), 402–421.

[van der Poorten 1979] A. van der Poorten, "A proof that Euler missed. . .Apéry's proof of the irrationality of $\zeta(3)$", *Math. Intelligencer* **1**:4 (1979), 195–203. MR 80i:10054 Zbl 0409.10028

[Silagadze 2010] Z. K. Silagadze, "Sums of generalized harmonic series for kids from five to fifteen", preprint, 2010. arXiv 1003.3602

Received: 2009-11-13 Accepted: 2010-06-29

jt17dava@siena.edu *Siena College, Department of Mathematics, 515 Loudon Road, Loudonville, NY 12211, United States*

nkrylov@siena.edu *Siena College, Department of Mathematics, 515 Loudon Road, Loudonville, NY 12211, United States*

INVOLVE 3:3(2010)

Infinite family of elliptic curves of rank at least 4

Bartosz Naskręcki

(Communicated by Bjorn Poonen)

We investigate $\mathbb{Q}$-ranks of the elliptic curve $E_t\colon\ y^2+txy=x^3+tx^2-x+1$, where t is a rational parameter. We prove that for infinitely many values of t the rank of $E_t(\mathbb{Q})$ is at least 4.

1. Introduction

In this paper we investigate the family of curves

$$E_t : y^2+txy=x^3+tx^2-x+1, \tag{1-1}$$

with parameter $t\in\mathbb{Q}$, and prove:

Main Theorem. *For infinitely many $u\in\mathbb{Q}$, the elliptic curve over $\mathbb{Q}$ given by the affine equation $E_{(u^2-u-3)} : y^2+(u^2-u-3)xy=x^3+(u^2-u-3)x^2-x+1$ has Mordell–Weil group of rank at least 4. More precisely, the group $E_{(u^2-u-3)}(\mathbb{Q})$ contains the subgroup spanned by the linearly independent points*

$$(0,1),\ (1,1),\ (u,u+1),\ \left(\tfrac{1}{9},\tfrac{1}{54}(9+3u-3u^2+v)\right),$$

where the point (u,v) lies on the elliptic curve given by the equation

$$2569+18u-9u^2-18u^3+9u^4=v^2.$$

The latter curve has Weierstrass model

$$y_0^2=x_0^3-92835x_0+1389150, \tag{1-2}$$

which defines an elliptic curve over $\mathbb{Q}$ with Mordell–Weil group of rank 2 and torsion group $\mathbb{Z}/2\mathbb{Z}$, spanned by the points

$$(-309,756),\ (-45,2340),\ (15,0).$$

MSC2000: 11D25, 11G05.

Keywords: elliptic curves, Mordell–Weil group, ranks in families.

This work was completed as a part of an undergraduate project at the Department of Mathematics and Computer Science in Poznań. We used the open source computer package Sage for almost all numerical computations.

The curves $E_{(u^2-u-3)}$ have different j-invariants for all but finitely many $u \in \mathbb{Q}$ as in the statement of the theorem.

Brown and Myers [2002] constructed an infinite family of elliptic curves over $\mathbb{Q}$ with quadratic growth of parameter and the rank of the Mordell–Weil group at least three. They asked whether one can find similar families of elliptic curves with higher ranks. Our Main Theorem resulted from attempts to answer the question. The method developed in this paper is modeled on the approach of [Brown and Myers 2002]. It naturally leads to computations with curves of high genera (instead of using specializations [Silverman 1986; 1983] as in the well-known method of Mestre).

It is of fundamental interest to find families of elliptic curves parametrized by a rational parameter with ranks higher than a prescribed constant [Kowalski 2007; Silverberg 2007; Rubin and Silverberg 2007; Kihara 1997; Mestre 1991; Nagao 1994]. The method of Mestre based on specialization theorems and computer search gives several infinite families of elliptic curves over $\mathbb{Q}$ of ranks as high as 14. Recent work by Elkies [2007] revealed elliptic curves with rank 18 over $\mathbb{Q}(t)$ and 19, parametrized by an elliptic curve of positive rank. Weierstrass equations of these families are rather complicated rational expressions of high order.

The family in Main Theorem provides quadratic polynomials as coefficients of the Weierstrass equation and four linearly independent points of a simple form. In addition, we obtain a general algorithm which can provide more such simple families with similar properties, and of rank at least 4. The main obstruction to obtaining higher ranks with our method is the base change from the projective line to a curve of higher genus — the best choice being a curve of genus 1 with infinite set of rational points over $\mathbb{Q}$ (as suggested after Lemma 2.2).

Rubin and Silverberg [2007] have obtained other infinite families of elliptic curves over $\mathbb{Q}$ of rank 4, constructed by twisting a curve given in the Legendre form. The families in that work are parametrized the by projective line or by an elliptic curve of rank 1 with twists parametrized by another elliptic curve of rank 1.

The choice of a particular family E_t was motivated by the study of more general families of elliptic curves with polynomial coefficients of degree at most one in the variable t. We first choose two rational constant sections with small coefficients. This method is likely to give rational elliptic surfaces with those two sections being independent. Then we look for a subfamily which contains a section with nonconstant x-coordinate, for example linear in variable t. Computations reveal what base change (e.g., quadratic base change from $\mathbb{P}^1$ to $\mathbb{P}^1$) shall increase the rank from 2 to 3 for particular values of t. Finally, we look for a suitable fourth point with constant x-coordinate. This provides a new base change to curve of higher genus (infinitely many curves of rank 4 occur only with elliptic curve with positive rank as a base). Similar computations gave us one more family of the type described in

Main Theorem, namely

$$F_{t(u)} : y^2 - t(u)xy = x^3 - t(u)x^2 - t(u)x + 1,$$

where $t(u) = 1 - u/2 + u^2/2$ and

$$v^2 = 361 + 198u - 189u^2 - 18u^3 + 9u^4$$

is the elliptic curve in a quartic form with rank 4 over $\mathbb{Q}$. For all but finitely many $u \in \mathbb{Q}$ the points

$$(0, 1),\ (2, 3),\ (u, u-1),\ \left(\tfrac{4}{9}, \tfrac{1}{27}(6 - 3u + 3u^2 + v)\right)$$

on the curve $F_{t(u)}$ are linearly independent.

The result stated in Main Theorem can be extended, if the parity conjecture holds true for E [Rohrlich 1994]. Let $\Lambda(E/\mathbb{Q}, s)$ be the complete L-series of the elliptic curve over $\mathbb{Q}$. Denote by $w(E) \in \{\pm 1\}$ the root number in the functional equation

$$\Lambda(E, 2-s) = w(E)\Lambda(E, s). \tag{1-3}$$

The parity conjecture predicts that

$$(-1)^{\text{rank } E(\mathbb{Q})} = w(E). \tag{1-4}$$

We can compute the root number $w(E_t)$ for the specific curves E_t and determine the parity of the rank of group $E_t(\mathbb{Q})$. Computations can be done explicitly using Sage [Stein et al. 2005], by choosing primes of bad reduction of E_t. We state numerical results in Section 4.2. In particular, assuming parity conjecture we constructed several elliptic curves over $\mathbb{Q}$ that have Mordell–Weil rank at least 5 (see Table 2).

2. Description of the algorithm

There are two obvious points lying on the curve (1-1), namely:

$$(0, 1), (1, 1) \in E_t(\mathbb{Q}(t)).$$

We produce with them several other points with coordinates in the ring $\mathbb{Z}[t]$:

$$\begin{aligned}
-(0, 1) &= (0, -1),\\
-(1, 1) &= (1, -t-1),\\
(0, 1) + 2(1, 1) &= (-t+1, -1),\\
(0, 1) + (1, 1) &= (-t-1, t^2+t-1),\\
(0, 1) - (1, 1) &= (t+3, 2t+5),\\
-(0, 1) + (1, 1) &= (t+3, -t^2-5t-5),\\
-(0, 1) + 2(1, 1) &= (t+5, 2t+11),\\
2(1, 1) &= (-1, t+1).
\end{aligned}$$

The following lemma describes the structure of the group $E_t(\mathbb{Q}(t))$.

Lemma 2.1. *The group $E_t(\mathbb{Q}(t))$ has rank 2 and has trivial torsion. It is generated by the points $(0, 1)$ and $(1, 1)$.*

Proof. Consider the elliptic curve

$$E_t : y^2 + txy = x^3 + tx^2 - x + 1$$

over $\mathbb{Q}(t)$ as the elliptic surface $\mathscr{E}$ over $\mathbb{P}^1$. The discriminant of E_t is equal to

$$-(t+2)^2(t^4+8t^3+11t^2-20t+92),$$

and the surface $\mathscr{E}$ has 6 singular fibers (in Kodaira classification):

- a fiber of type I_6 over $t = \infty$,
- a fiber of type I_2 over $t = -2$,
- four fibers of type I_1 over $t = \alpha_i$ for $i = 1, 2, 3, 4$, where

$$t^4+8t^3+11t^2-20t+92 = (t-\alpha_1)(t-\alpha_2)(t-\alpha_3)(t-\alpha_4).$$

Let S be the set of bad places, namely $S = \{\infty, -2, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. From the Shioda–Tate formula [Shioda 1990, Corollary 5.3] we get

$$\operatorname{rank} E_t(\overline{\mathbb{Q}}(t)) = \rho(\mathscr{E}) - 2 - \sum_{v \in S} (m_v - 1),$$

where $\rho(\mathscr{E})$ is the Picard number of surface $\mathscr{E}$ and m_v is the number of components of singular fiber over place v. After [Shioda 1990, Equation 10.14] we find that the elliptic surface $\mathscr{E}$ is rational since the coefficients of the defining equation in Weierstrass form satisfy the condition:

$$\deg a_i(t) \le i,$$

and the discriminant is nonconstant. This implies that $\rho(\mathscr{E}) = 10$ [Shioda 1990, Lemma 10.1] and we get from the Shioda–Tate formula:

$$\operatorname{rank} E_t(\overline{\mathbb{Q}}(t)) = 2.$$

Computation of the height pairing matrix for the points $P_1 = (0, 1)$ and $P_2 = (1, 1)$ gives the matrix:

$$(\langle P_i, P_j \rangle)_{1 \le i,j \le 2} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{6} \end{pmatrix}.$$

This shows that points P_1 and P_2 span the free part of the group $E_t(\overline{\mathbb{Q}}(t))$. Since they are both rational points over $\mathbb{Q}(t)$, it follows that they span the free part of the group $E_t(\mathbb{Q}(t))$.

The map

$$\phi : E_t(\overline{\mathbb{Q}}(t)) \to \prod_{v \in S} G(F_v)$$

takes a section to the respective fiber component of F_v that it meets. The group $G(F_v)$ is generated by simple components of the fiber F_v. The map ϕ is an injection on the torsion part. From the Néron model structure we know that for the multiplicative fibers I_n the group $G(I_n) \cong \mathbb{Z}/n\mathbb{Z}$. In case of the family E_t we get the injection

$$E_t(\overline{\mathbb{Q}}(t))_{\text{tors}} \hookrightarrow \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}.$$

Let $P = (x, y) \in E_t(\overline{\mathbb{Q}}(t))$ be a 2-torsion point. The condition $P = -P$ implies that x-coordinate must satisfy:

$$0 = 4 - 4x + 4tx^2 + t^2x^2 + 4x^3.$$

The polynomial on the right side is irreducible over $\overline{\mathbb{Q}}(t)$. Similarly, let P satisfy $P = -2P$. It follows that

$$0 = -1 + 4t + t^2 + 12x - 6x^2 + 4tx^3 + t^2x^3 + 3x^4.$$

Again the polynomial is irreducible over $\overline{\mathbb{Q}}(t)$. This clearly implies that only the point at infinity has finite order:

$$E_t(\overline{\mathbb{Q}}(t))_{\text{tors}} = \{\mathbb{O}\}. \qquad \square$$

In order to find more points on the curve (1-1) we specialize parameter t to a polynomial function of another parameter u:

$$t(u) = a_n u^n + \cdots + a_1 u + a_0,$$

where $a_i \in \mathbb{Q}$. To get a rational point on the curve (1-1) with x-coordinate equal to $au + b$, for $a, b \in \mathbb{Q}$, it is necessary and sufficient that

$$\Delta(u) = 4(-1 + b + au)^2(1 + b + au) + (2 + t(u))^2(b + au)^2 \qquad (2\text{-}1)$$

be a perfect square.

Lemma 2.2. *Let $P = (x, y)$ be a rational point on the curve $E_{t(u)}$ over $\mathbb{Q}(u)$, where $t(u) = a_n u^n + \cdots + a_1 u + a_0 \in \mathbb{Q}[u]$ of positive degree. Let $x = au + b \in \mathbb{Q}[u]$. Suppose that $a \neq 0$.*

(i) *If $\deg t = 1$, then $P = k(0, 1) + l(1, 1)$ for some $k, l \in \mathbb{Z}$.*

(ii) *If $\deg t = 2$, then $P = (x, x + 1)$ and $t(u) = x^2 - x - 3$ or $P = (x, x - 1)$ and $t(u) = -x^2 + x + 1$.*

If, in addition, $\deg t > 2$, there is no rational point whose x-coordinate is equal to $au + b \in \mathbb{Q}[u]$.

Proof. (i) Assume $t(u)=a_1u+a_0$ and $a_1\neq 0$. Put $P(u)=q_2u^2+q_1u+q_0$. Since P is a rational point on $E_{t(u)}$, the discriminant $\Delta(u)$ as in (2-1) is a perfect square:

$$\Delta(u)=P(u)^2;$$

moreover, for some $\varepsilon\in\{1,-1\}$, we have $q_2=\varepsilon a a_1$, $q_1=\dfrac{2a^2+2aa_1+aa_0a_1+ba_1^2}{\varepsilon a_1}$, and

$$q_0=\frac{-2a^3-4a^2a_1-2a^2a_0a_1-2aa_1^2+4aba_1^2+2ba_1^3+ba_0a_1^3}{\varepsilon a_1^3}.$$

Equating the last two coefficients of $\Delta(u)$ and $P(u)^2$ gives two equations in the variables a_1, a_0, a, b:

$$\begin{aligned}R_1(a,b,a_0,a_1)=&-a^6-4a^5a_1-2a^5a_0a_1-6a^4a_1^2+4a^4ba_1^2-4a^4a_0a_1^2-a^4a_0^2a_1^2\\&-4a^3a_1^3+10a^3ba_1^3-2a^3a_0a_1^3+5a^3ba_0a_1^3-a^2a_1^4+8a^2ba_1^4\\&-4a^2b^2a_1^4+4a^2ba_0a_1^4+a^2ba_0^2a_1^4+2aba_1^5-4ab^2a_1^5+aba_0a_1^5\\&-2ab^2a_0a_1^5+a_1^6-ba_1^6-b^2a_1^6+b^3a_1^6\\=&\,0,\end{aligned}$$

$$\begin{aligned}R_2(a,b,a_0,a_1)=&\,2a^4+6a^3a_1+3a^3a_0a_1+6a^2a_1^2-3a^2ba_1^2+4a^2a_0a_1^2+a^2a_0^2a_1^2\\&+2aa_1^3-4aba_1^3+aa_0a_1^3-2aba_0a_1^3-a_1^4-ba_1^4+b^2a_1^4\\=&\,0.\end{aligned}$$

The ideal $I=I(R_1,R_2)$ of these equations can be rearranged in the form of the Gröbner basis $I=I(a^9-2a^7a_1^2+a^5a_1^4, R_1', \ldots, R_{18}')$, with $R_i'=R_i'(a,b,a_0,a_1)$. The first polynomial of the new basis factors as $a^5(a-a_1)^2(a+a_1)^2$. The equation $a^5(a-a_1)^2(a+a_1)^2=0$ can only have solutions $a=\pm a_1$ since we assumed $a\neq 0$. For $a=a_1$ the equations reduce to

$$a_0=b-3 \quad \text{or} \quad a_0=b-5.$$

For $t(u)=au+(b-3)$ or $t(u)=au+(b-5)$, we get respectively the points

$$\begin{gathered}\big(t(u)+3,\,5+2t(u)\big),\ \big(t(u)+3,\,-5-5t(u)-t(u)^2\big),\\ \big(t(u)+5,\,11+2t(u)\big),\ \big(t(u)+5,\,-11-7t(u)-t(u)^2\big).\end{gathered}$$

They are linear combinations of $(0,1)$ and $(1,1)$ in the group $E_t(\mathbb{Q}(u))$.

For $a=-a_1$ the equations reduce to

$$a_0=-1-b \quad \text{or} \quad a_0=1-b.$$

For $t(u) = -au - b - 1$ and $t(u) = -au - b + 1$ we get the points

$$\big(-t(u)-1, 1\big),\ \big(-t(u)-1, -1+t(u)+t(u)^2\big),$$
$$\big(-t(u)+1, -1\big),\ \big(-t(u)+1, 1-t(u)+t(u)^2\big),$$

respectively. Again both points are the linear combinations of (0, 1) and (1, 1).

We can now proceed analogously to the proof of (i) and show property (ii). Let $t(u) = a_2u^2 + a_1u + a_0$ and $a_1 \neq 0$. Put $P(u) = q_3u^3 + q_2u^2 + q_1u + q_0$. Comparing coefficients of (2-1) and $P(u)^2$ implies

$$q_3 = \varepsilon a a_2, \quad q_2 = \frac{aa_1 + ba_2}{\varepsilon}, \quad q_1 = \frac{2a + aa_0 + ba_1}{\varepsilon}, \quad q_0 = \frac{2a^2 + 2ba_2 + ba_0a_2}{\varepsilon a_2},$$

with $\varepsilon = \pm 1$. By comparing the three lowest terms in $P(u)^2$ and $\Delta(u)$ we obtain

$$a_1 = \frac{(-1+2b)a_2}{a}, \quad a^3(2+a_0) + a(1+b-b^2)a_2 = 0, \quad a^2 = \lambda a_2,$$

with $\lambda = \pm 1$. This implies that $t(u) = -2 - \lambda - (au+b)\lambda + (au+b)^2\lambda$. In this way — assuming $x(u) = au + b$ — we get the two distinct families

	point	parameter $t(u)$
Family A	$(x, x+1)$	$x^2 - x - 3$
Family B	$(x, x-1)$	$-x^2 + x + 1$

To show the last case we proceed by induction on degree of the polynomial $t(u)$. Consider $t(u)$ as a polynomial in u of degree $n > 2$; then $\deg \Delta = 2n + 2$, so we look for the polynomial $P(u)$ of degree $n+1$ such that $\Delta(u) = P(u)^2$. We put

$$a_i^* = \begin{cases} a_i & \text{if } 0 < i \leq n, \\ a_0 + 2 & \text{if } i = 0, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad q_j^* = \begin{cases} q_j & \text{if } 0 \leq j \leq n+1, \\ 0 & \text{otherwise.} \end{cases}$$

We prove by induction the formula

$$q_j^* = \varepsilon(aa_{j-1}^* + bb_j^*),$$

using the identities

$$c_j(\Delta) = a^2 \sum_{j=\alpha+\beta} a_\alpha^* a_\beta^* + 2ab \sum_{j+1=\alpha+\beta} a_\alpha^* a_\beta^* + b^2 \sum_{j+2=\alpha+\beta} a_\alpha^* a_\beta^*, \quad c_j(P^2) = \sum_{j=\alpha+\beta} q_\alpha^* q_\beta^*,$$

for $j = n+1, \ldots, 2n+2$, where $c_j(a_0 + a_1x + \cdots + a_nx^n) = a_j$. It follows from $\Delta = P^2$ that

$$c_j(\Delta) = c_j(P^2).$$

We substitute the coefficients $q_0 = \varepsilon(2b + ba_0)$ and $q_1 = \varepsilon(a(a_0+2) + ba_1)$ into the identities above with $j = 0, 1, 2$ and we get $b^2 = 1$ and finally $a = 0$, a contradiction. This completes the proof of the lemma. □

By Lemma 2.2, we can specialize to one of the quadratic parameters, since the families A and B (see previous page) have similar properties. For the rest of the paper we choose the specialization $t(u) = u^2 - u - 3$:

$$E_{(u^2-u-3)} : y^2 + (u^2-u-3)xy = x^3 + (u^2-u-3)x^2 - x + 1.$$

For notational simplicity, we write $f(x, y, u) = 0$ for this equation. The point $(u, u+1)$ lies on these curves and gives several new integral points over $\mathbb{Q}[u]$:

$$\begin{aligned}
-(u, u+1) &= (u, -u^3 + u^2 + 2u - 1),\\
(0, 1) + (u, u+1) &= (-u+1, u^3 - 2u^2 - u + 1),\\
(1, 1) - (u, u+1) &= (u^3 - 2u, u^4 + u^3 - 3u^2 - 2u + 1),\\
2(1, 1) + (u, u+1) &= (-u^3 + 4u^2 - 6u + 4, u^5 - 6u^4 + 14u^3 - 17u^2 + 10u - 1).
\end{aligned}$$

To find the fourth linearly independent rational point on the curve $E_{(u^2-u-3)}$, we consider the following general algorithm:

(1) Choose two rational functions $a(x), b(x) \in \mathbb{Q}(x)$.

(2) Form the simultaneous equations $f(a(u), y_a(u), u) = 0$, $f(b(u), y_b(u), u) = 0$.

(3) Find $a(x), b(x)$ such that $y_a(x), y_b(x) \in \mathbb{Q}(x)$.

(4) A sufficient and necessary condition for y_a, y_b to be rational is that the discriminant of the quadratic equation $f(a(x), y_a, x) = 0$ in y_a be a perfect square. The same condition holds for the equation in y_b.

(5) Find all rational points, i.e., the triples $(u, s, w) \in \mathbb{Q}^3$ on the affine curve:

$$\Delta_{f(a(x),y_a,x)=0}(u) = s^2, \qquad \Delta_{f(b(x),y_b,x)=0}(u) = w^2, \tag{2-2}$$

where $\Delta_{f(a(x),y_a,x)=0}(x)$ and $\Delta_{f(b(x),y_b,x)=0}(x)$ belong to $\mathbb{Q}(x)$.

We now pick $a(x) = x$ and $b(x) = c$. Then the first equation in (2-2) reduces to $(2 - u - u^2 + u^3)^2 = s^2$, while the second gives

$$4 - 4c - 3c^2 + 4c^3 + 2c^2u - c^2u^2 - 2c^2u^3 + c^2u^4 = w^2.$$

We choose $c \in \mathbb{Q}$ so that it defines the elliptic curve in a quartic form with infinitely many points (u, w). A direct search with $u \in \mathbb{N}$ reveals that for $u = 7$ we have on the curve E_{39} the four linearly independent points

$$(0, 1),\ (1, 1),\ (7, 8),\ \left(\tfrac{1}{9}, \tfrac{8}{27}\right);$$

hence we put $c = \frac{1}{9}$, as in the statement of Main Theorem.

3. Proofs

To prove Main Theorem, we will need the following elementary lemma.

Lemma 3.1. *Let $b \in M$, where M is a left $\mathbb{Z}$-module. Suppose $a_1, \ldots, a_k \in M$ are linearly independent over $\mathbb{Z}$ and the nonzero cosets $[a_1], \ldots, [a_k] \in M/2M$ are linearly independent over $\mathbb{F}_2$. If $[b] \notin \langle [a_1], \ldots, [a_k] \rangle$ and the 2-torsion of M is trivial, then $b, a_1, \ldots, a_k$ are independent over $\mathbb{Z}$ in M.*

Proof. Suppose, contrary to our claim, that there exists $\alpha_1, \ldots, \alpha_k, \beta \in \mathbb{Z}$, not all zero, such that $\beta b + \alpha_1 a_1 + \cdots + \alpha_k a_k = 0$. We can assume that β is the least positive integer for which this holds. If β is odd, we have $[\beta b] = [b]$ and $[b] = [\alpha_1 a_1 + \cdots + \alpha_k a_k]$, a contradiction. If β is even, we have $[0] = [\alpha_1 a_1 + \cdots + \alpha_k a_k]$; but the linear independence of cosets $[a_i]$ over $\mathbb{F}_2$ implies that all α_i are even, so it is possible to write $\beta' b = \alpha_1' a_1 + \cdots + \alpha_k' a_k$, where $2\beta' = \beta$ and $2\alpha_i' = \alpha_i$. This contradicts the minimality of β. □

We now establish the structure of the torsion subgroup of the curve E_t for all but finitely many $t \in \mathbb{Q}$.

Lemma 3.2. *Let $t_1(u) = u$ and $t_2(u) = u^2 - u - 3$. The structure of the torsion subgroup of groups $E_{t_i(u)}(\mathbb{Q})$ for $u \in \mathbb{Q}$ is as follows*:

Group T	$\#\{u \in \mathbb{Q} : E_{t_1(u)}(\mathbb{Q})_{\text{tors}} \cong T\}$	$\#\{u \in \mathbb{Q} : E_{t_2(u)}(\mathbb{Q})_{\text{tors}} \cong T\}$
$\mathbb{Z}/2\mathbb{Z}$	∞	0
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2N\mathbb{Z}$ $N = 1, 2, 3, 4$	0	0
$\mathbb{Z}/4\mathbb{Z}$, $\mathbb{Z}/8\mathbb{Z}$	$< \infty$	0
$\mathbb{Z}/3N\mathbb{Z}$ $N = 1, 2, 3, 4$	0	0
$\mathbb{Z}/5\mathbb{Z}$, $\mathbb{Z}/10\mathbb{Z}$	$< \infty$	$< \infty$
$\mathbb{Z}/7\mathbb{Z}$	$< \infty$	$< \infty$

Proof. Mazur [Mazur 1978] showed that the group $E_t(\mathbb{Q})_{\text{tors}}$ is isomorphic either to $\mathbb{Z}/N\mathbb{Z}$ with $1 \le N \le 10$ or $N = 12$, or to $\mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ with $N = 2, 4, 6$ or 8. We prove below that for $t = u^2 - u - 3$, with $u \in \mathbb{Q}$, the groups $E_t(\mathbb{Q})_{\text{tors}}[2]$ and $E_t(\mathbb{Q})_{\text{tors}}[3]$ are trivial for all u. The triviality of rational 5-torsion and 7-torsion subgroups is proved only for all but finitely many u. The 2-torsion is computed also for the general parameter t, which leads to curve of genus zero with rational parametrization. These facts, combined with Mazur's theorem, will suffice to finish the proof of the lemma.

Let $P=(x,y)$ be a 2-torsion point on the curve $y^2+txy=x^3+tx^2-x+1$. Its negative is $-P=(x,-tx-y)$. From the condition $P=-P$ it follows that

$$y=-\frac{tx}{2}.$$

Substitution into the Weierstrass equation of E_t gives the equation

$$0=4-4x+4tx^2+t^2x^2+4x^3, \tag{3-1}$$

which defines the curve of genus zero with the parametrization

$$t=\frac{-8-8s-2s^2-s^3}{4+s^2}, \quad x=-1-\frac{s^2}{4}.$$

When this is substituted in (3-1), the only nontrivial 2-torsion point is obtained:

$$\left(-1-\tfrac{1}{4}s^2,\ \tfrac{1}{8}(-8-8s-2s^2-s^3)\right).$$

Therefore the groups $\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2N\mathbb{Z}$ for $N=1,2,3,4$ cannot occur as torsion subgroups of $E_t(\mathbb{Q})$. Specialization of the parameter $t(u)=u^2-u-3$ gives the curve of genus two

$$C_1: 4-4x+4(-3-u+u^2)x^2+(-3-u+u^2)^2x^2+4x^3=0,$$

which has the normal form

$$C_2: Y^2=(-5-2X-X^2)(11+34X+7X^2+4X^3).$$

where

$$X=\frac{1+ux-u^2x}{x-1}, \quad Y=-4x+8ux. \tag{3-2}$$

We define $\mathrm{Jac}(C_2)$ to be the Jacobian variety of the curve C_2 over $\mathbb{Q}$. The group $\mathrm{Jac}(C_2)(\mathbb{Q})$ of rational points of this variety has the torsion subgroup isomorphic to $\mathbb{Z}/4\mathbb{Z}$ and 2-Selmer group $\mathrm{Sel}(\mathrm{Jac}(C_2)/\mathbb{Q})[2]$ over $\mathbb{Q}$ isomorphic to $\mathbb{Z}/2\mathbb{Z}$ (computed with the Magma commands TorsionSubgroup and TwoSelmerGroup). This enables us to perform a two-descent and compute that the rank $\mathrm{Jac}(C_2)(\mathbb{Q})=0$.

The only rational points on the curve C_2 might come from the torsion points of the Jacobian. We compute (with the Magma Chabauty0 procedure) that actually only the point at infinity is the rational point on C_2. The affine points on the curve C_1 come from the affine part of C_2 via the map defined in (3-2), except for the points $(x,u)=(1,\frac{1}{2}(1\pm\sqrt{5}))$; hence there are only two rational points on C_1: the points at infinity (the parameter $u=\infty$ defines a singular curve). This shows that $E_{u^2-u-3}(\mathbb{Q})_{\mathrm{tors}}[2]=\{\mathbb{O}\}$ for all $u\in\mathbb{Q}$ such that E_{u^2-u-3} is nonsingular.

Let $P=(x,y)$ be a 3-torsion point on the curve $y^2+txy=x^3+tx^2-x+1$. The condition $P=-2P$ implies that the pair (x,t) must satisfy the equation

$$C_3: 1-4t-t^2-12x+6x^2-4tx^3-t^2x^3-3x^4=0.$$

The curve C_3 has genus two and has the normal form

$$C_4 : Y^2 = (1-X)(5+3X)(1+X^3),$$

where

$$X = x, \quad Y = \frac{2+t+2x^3+tx^3}{1-x}.$$

Similarly to the case of 2-torsion we compute that the rank of $\mathrm{Jac}(C_4)(\mathbb{Q})$ is zero because $\mathrm{Sel}(\mathrm{Jac}(C_4)/\mathbb{Q})[2] \cong \mathbb{Z}/8\mathbb{Z}$ and $\mathrm{Jac}(C_4)(\mathbb{Q})_{\mathrm{tors}} \cong \mathbb{Z}/8\mathbb{Z}$. We obtain four rational points on C_3: two at infinity, plus

$$(x,t) = \left(-\tfrac{5}{3}, -2\right) \quad \text{and} \quad (x,t) = (1,-2).$$

The parameter $t = -2$ gives a singular nodal curve; hence $E_t(\mathbb{Q})_{\mathrm{tors}}[3] = \{\mathbb{O}\}$, for all $t \in \mathbb{Q}$ such that E_t is nonsingular.

Let $P = (x, y) \in E_t(\mathbb{Q})$ be the point of order 4. The conditions $2P = -2P$ and $2P \neq \mathbb{O}$ imply that the pair (x, t) satisfies the equation

$$\begin{aligned}-14-8t-2t^2+8x-4tx+15t^2x+8t^3x+t^4x-10x^2+40tx^2+10t^2x^2\\ +40x^3-10x^4+4tx^5+t^2x^5+2x^6=0,\end{aligned}$$

which defines a curve of genus 3. From Faltings' theorem [1983] we see that for all but finitely many $t \in \mathbb{Q}$ groups $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/8\mathbb{Z}$ cannot occur as torsion subgroups of $E_t(\mathbb{Q})$.

The cases of rational 5-torsion and 7-torsion do not generate hyperelliptic curves so we use again Faltings' theorem.

The condition $3P = -2P$ for $P = (x, y) \in E_t(\mathbb{Q})$ implies that the pair (x, t) lies on the curve of genus 13 given by

$$\begin{aligned}&223+140t-13t^2+40t^3+45t^4+12t^5+t^6-540x-480tx+200t^2x+160t^3x+20t^4x\\
&\quad+190x^2+1680tx^2+900t^2x^2+240t^3x^2+30t^4x^2+1520x^3-820tx^3-685t^2x^3\\
&\quad-560t^3x^3-270t^4x^3-60t^5x^3-5t^6x^3-1795x^4-120tx^4-430t^2x^4+440t^3x^4+455t^4x^4\\
&\quad+120t^5x^4+10t^6x^4+696x^5+112tx^5+1372t^2x^5+736t^3x^5-124t^4x^5-244t^5x^5\\
&\quad-95t^6x^5-16t^7x^5-t^8x^5-60x^6+720tx^6+420t^2x^6-840t^3x^6-705t^4x^6-180t^5x^6\\
&\quad-15t^6x^6+240x^7+360tx^7-1350t^2x^7-720t^3x^7-90t^4x^7+105x^8-1140tx^8-285t^2x^8\\
&\quad-380x^9+80tx^9+20t^2x^9+62x^{10}-16t^2x^{10}-8t^3x^{10}-t^4x^{10}-20tx^{11}-5t^2x^{11}-5x^{12}\\
&=0.\end{aligned}$$

The condition $4P = -3P$ for $P = (x, y) \in E_t(\mathbb{Q})$ implies that the pair (x, t) lies on a curve of genus 31 (equation omitted).

Applying the theorem of Faltings, we deduce that, for all but finitely many t, the rational 5-torsion and 7-torsion are trivial. Mazur's structure theorem now implies the statement of the lemma. □

Proof of Main Theorem. In order to prove Main Theorem we check if the appropriate points and their linear combinations belong to $2E_{(u^2-u-3)}(\mathbb{Q})$. Given a $\mathbb{Q}$-rational point $P=(x, y)$ on the curve $E_{(u^2-u-3)}$ over $\mathbb{Q}$ we have the following formula for the x-coordinate of the point $2P$:

$$x(2P)=\frac{4-2u+u^2+2u^3-u^4-8x+2x^2+x^4}{4-4x+(-3+2u-u^2-2u^3+u^4)x^2+4x^3}.$$

To simplify the notation, define:

$$P_{\varepsilon_1,\varepsilon_2,\varepsilon_3}=\varepsilon_1(0, 1)+\varepsilon_2(1, 1)+\varepsilon_3(u, u+1).$$

If for $u\in\mathbb{Q}$ there exists a rational point $\left(\frac{1}{9}, y\right)$ on the curve $E_{(u^2-u-3)}$ and y is one of two possible values

$$y=\tfrac{1}{54}\left(9+3u-3u^2\pm\sqrt{2569+18u-9u^2-18u^3+9u^4}\right),$$

then we put

$$Q_{\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4}=\varepsilon_1(0, 1)+\varepsilon_2(1, 1)+\varepsilon_3(u, u+1)+\varepsilon_4\left(\tfrac{1}{9}, y\right), \tag{3-3}$$

where $\varepsilon_i\in\{-1, 0, 1\}$.

The proof falls naturally into two parts. In the first part we establish the criteria for which the equations $P_{\varepsilon_1,\varepsilon_2,\varepsilon_3}=x(2P)$ and $Q_{\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4}=x(2P)$ have solutions in pairs of rational numbers (u, x) (recall that $P=(x, y)$ lies on $E_{(u^2-u-3)}$). In the second part of the proof we gather information to find the infinite subset of $\mathbb{Q}$ of parameters u for which the rank is at least 4. To use Lemma 3.1 we must consider the tuples $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ in

$$\{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1),\ (1, 1, 0),\ (0, 1, -1),\ (1, 0, 1),\ (1, 1, 1)\}. \tag{3-4}$$

Assume that $\left(\frac{1}{9}, y\right)$ is $\mathbb{Q}$-rational. Consider $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$ in the set

$$\{(0, 0, 0, 1),\ (1, 0, 0, 1),\ (0, 1, 0, 1),\ (0, 0, 1, 1),\ (1, 1, 0, 1),\ (0, 1, -1, 1),\ (1, 0, 1, 1),\ (1, 1, 1, 1)\}.$$

The tuples with negative entries were chosen to lower the genera of corresponding curves. Since we work mod $2E_{(u^2-u-3)}(\mathbb{Q})$ the tuples can be chosen with a fair amount of freedom. We compute genera of curves using the *genus* command from the *algcurves* package in Maple 12. We consider in detail three specific cases.

$(\varepsilon_1, \varepsilon_2, \varepsilon_3)=(1, 0, 0)$. This tuple implies the equation

$$\frac{4-2u+u^2+2u^3-u^4-8x+2x^2+x^4}{4-4x+(-3+2u-u^2-2u^3+u^4)x^2+4x^3}=0. \tag{3-5}$$

Since (0, 1) is not the point at infinity, the denominator is nonvanishing and

$$4-2u+u^2+2u^3-u^4-8x+2x^2+x^4=0.$$

It defines an elliptic curve of rank 1. By means of the formulas

$$x_0=\frac{1}{3(u^2-u-1)}(12u^4-12u^3x-36u^3+12u^2x^2+30u^2x+35u^2 \\ -12ux^3-24ux^2-42ux+61u+18x^3+6x^2+36x-113),$$

$$y_0=-\frac{2}{u^2-u-1}(8u^5-8u^4x-32u^4+8u^3x^2+28u^3x+44u^3-8u^2x^3-24u^2x^2 \\ -39u^2x+9u^2+20ux^3+20ux^2+43ux-101u-19x^3-11x^2-54x+122),$$

we can transform the equation into short Weierstrass form:

$$y_0^2=x_0^3+\tfrac{359}{3}x_0+\tfrac{3130}{27}.$$

The Mordell–Weil group of this elliptic curve is generated by the point $\left(\frac{53}{3}, 88\right)$. Hence in the original form the generator is equal to $(u, x)=\left(\frac{1}{2}, \frac{1}{2}\right)$. The remaining cases for $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ in (3-4) are summarized in the first table below; the genera were computed with Maple.

ε_1	ε_2	ε_3	genus
1	0	0	1
0	1	0	3
0	0	1	2
1	1	0	3
0	1	−1	4
1	0	1	2
1	1	1	4

ε_1	ε_2	ε_3	ε_4	genus
0	0	0	1	5
1	0	0	1	9
0	1	0	1	13
0	0	1	1	19
1	1	0	1	13
0	1	−1	1	15
1	0	1	1	11
1	1	1	1	28

Now assume that we are given a rational point $\left(\frac{1}{9}, y\right)$ lying on the curve $E_{(u^2-u-3)}$ for a suitable $u \in \mathbb{Q}$.

$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (0, 0, 0, 1)$. In this case

$$\frac{-u^4+2u^3+u^2-2u+x^4+2x^2-8x+4}{(u^4-2u^3-u^2+2u-3)x^2+4x^3-4x+4}=\tfrac{1}{9}. \tag{3-6}$$

This equation defines an affine curve of genus 5.

$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (1, 0, 0, 1)$. Here we have

$$\frac{-u^4+2u^3+u^2-2u+x^4+2x^2-8x+4}{(u^4-2u^3-u^2+2u-3)x^2+4x^3-4x+4} = -9u^2+9u-162y+180. \quad (3\text{-}7)$$

From the equation of the curve $E_{(u^2-u-3)}$ we can find the formula

$$y = \tfrac{1}{54}(-3u^2+3u+v+9), \quad (3\text{-}8)$$

with $v = \pm\sqrt{2569+18u-9u^2-18u^3+9u^4}$. Using these relations we can assume that if a point (u, x) lies on the curve given by (3-7), it also lies on the curve

$$9(9u^4-18u^3-9u^2+18u+2569)(u^4x^2-2u^3x^2-u^2x^2+2ux^2+4x^3-3x^2-4x+4)^2$$
$$-(153u^4x^2+u^4-306u^3x^2-2u^3-153u^2x^2-u^2+306ux^2+2u-x^4$$
$$+612x^3-461x^2-604x+608)^2 = 0.$$

This curve has genus 9. The rest is computed in a similar way; the results are given in the second table on the previous page.

In the last step of the proof we show for which $u \in \mathbb{Q}$ the point $\left(\frac{1}{9}, y\right)$ is $\mathbb{Q}$-rational. By formula (3-8) $y \in \mathbb{Q}$ if and only if $2569+18u-9u^2-18u^3+9u^4$ is a full square. This condition defines the elliptic curve in a quartic form. It is birational to elliptic curve in the Weierstrass form:

$$y_0^2 = x_0^3 - 92835x_0 + 1389150. \quad (3\text{-}9)$$

The Mordell–Weil group of the curve has rank 2. The torsion subgroup is isomorphic to $\mathbb{Z}/2$. Generators of the free part are $(x_0, y_0) = (-309, -756)$, and $(x_0, y_0) = (390, -4950)$ and the generator of the torsion subgroup is $(15, 0)$. In the quartic form they correspond respectively to

$$\left(\tfrac{1}{9}, \tfrac{1369}{27}\right),\ \left(\tfrac{27}{10}, -\tfrac{5173}{100}\right),\ (-6, -133).$$

The birational map between models of elliptic curve provides a method to generate a suitable infinite set S of parameters $u \in \mathbb{Q}$ (see Section 4.2 for details). In fact $\{u \in \mathbb{Q} : \text{rank}\,(E_{(u^2-u-3)}(\mathbb{Q})) \geq 4\} \subset S$ and the difference between sets correspond precisely to the set of u-coordinates of rational points on the curves listed in the tables on the previous page. Except for the case $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (1, 0, 0)$, all curves have finitely many rational points due to Faltings' theorem. Consider the curve $E_t : y^2+txy = x^3+tx^2-x+1$, and assume the point $\left(\frac{1}{9}, y\right)$ on the curve E_t is $\mathbb{Q}$-rational. Solving the quadratic equation gives

$$y = \tfrac{1}{54}\left(-3t \pm \sqrt{2596+36t+9t^2}\right).$$

The point $(0, 1)$ is a double in $E_t(\mathbb{Q})$ and $\left(\frac{1}{9}, y\right)$ is a rational point when there exists a triple $(t, x, s) \in \mathbb{Q}^3$ on

$$s^2 = 2596 + 36t + 9t^2, \quad 0 = 1 - 4t - t^2 - 8x + 2x^2 + x^4.$$

The parametrization of the first equation gives

$$t = \frac{2596 - f^2}{6f - 36}, \qquad s = \frac{f^2 - 12t + 2596}{2f - 12},$$

with a new parameter $f \in \mathbb{Q}\backslash\{6\}$. Substituting into the second equation gives the curve:

$$\begin{aligned} -6364096 - 62736f + 5084f^2 + 24f^3 - f^4 - 10368x + 3456fx - 288f^2x \\ + 2592x^2 - 864fx^2 + 72f^2x^2 + 1296x^4 - 432fx^4 + 36f^2x^4 = 0. \end{aligned}$$

This curve has genus 3, so it has finitely many rational points by Faltings' theorem. Specializing to a parameter $t(u) = u^2 - u - 3$ we obtain that there are only finitely many $u \in \mathbb{Q}$ for which $(0, 1)$ is a double in $E_{(u^2-u-3)}(\mathbb{Q})$ while $\left(\frac{1}{9}, y\right)$ is $\mathbb{Q}$-rational. So in fact the difference $S\backslash B$ is a finite set.

It remains to show that the j-invariant of the curve $E_t : y^2 + txy = x^3 + tx^2 - x + 1$ repeats itself for finitely many $t \in \mathbb{Q}$. We compute

$$j(E_t) = -\frac{(48 + t^2(4 + t)^2)^3}{(2 + t)^2(92 + (-1 + t)t(4 + t)(5 + t))}. \tag{3-10}$$

Hence the equation

$$j(E_{t_1}) = j(E_{t_2})$$

defines an affine curve with coordinates (t_1, t_2) which has genus 11 according to computations in Maple. This implies that specializing the parameter t to $u^2 - u - 3$ gives a curve with finitely many rational points. □

4. Numerical results

4.1. ***General statistics.*** We show in Figure 1, left, the rank of the curves

$$y^2 + txy = x^3 + tx^2 - x + 1, \tag{4-1}$$

with positive integers $t < 230$. All computations were performed with Sage 3.4 [Stein et al. 2005] using the `mwrank` procedure. In some 6% of cases the value shown is conjectural, since it was not possible to prove an upper bound for the rank. Here is the percentage of curves of each rank:

rank	1	2	3	4	?
fraction	1%	41%	45%	7%	6%

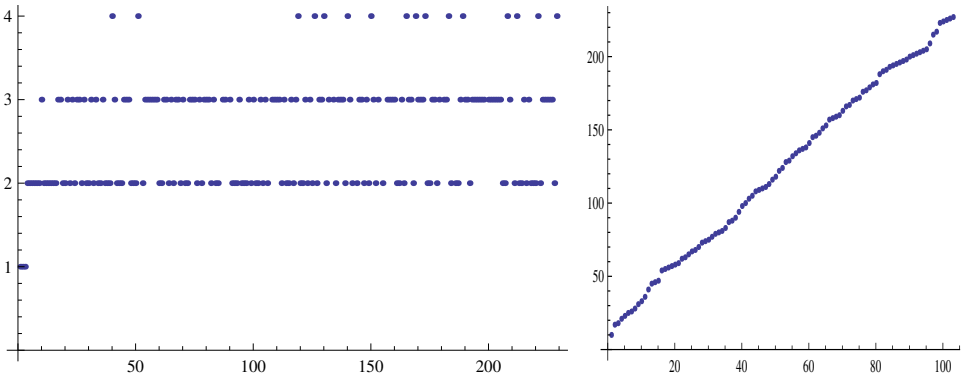


Figure 1. Left: Rank of curves of the form (4-1). Right: Growth of t for curves of rank 3 (abscissa: number of curves of rank 3 up to a certain t).

Also interesting is the plot of curves of rank 3 in Figure 1, right. It suggests that the progression for curves of rank 3 is almost linear, and hence that we can use the general algorithm from the introduction to state another version of the main Main Theorem and find many more infinite families of elliptic curves over $\mathbb{Q}$.

4.2. ***Explicit version of the main theorem.*** The statement of the theorem requires removing a finite subset of "bad rational points" due to Faltings' theorem (Mordell's conjecture). The upper bound of heights of this points is hard to obtain. We shall give an explicit and effective version of the main result of this paper. For rational points on the curve (1-2) with low height we can compute the explicit table of corresponding elliptic curves $E_{(u^2-u-3)}$ over $\mathbb{Q}$ of rank at least 4. The curve

$$C_1 : y_0^2 = x_0^3 - 92835x_0 + 1389150 \tag{4-2}$$

is mapped to the curve

$$C_2 : 2569 + 18u - 9u^2 - 18u^3 + 9u^4 = v^2 \tag{4-3}$$

via the map

$$\phi : C_1 \to C_2,$$

where $\phi(x_0, y_0) = (u, v)$ is given by the formulas

$$u = \frac{565605 + x_0(-948 + 7x_0) + 266y_0}{(-1551 + x_0)(45 + x_0)},$$

$$v = \frac{1}{(-1551 + x_0)^2(45 + x_0)^2}\big(133(92385 + (-30 + x_0)x_0)(-115425 + x_0(1536 + x_0)) + 234(-3922935 + x_0(9010 + 41x_0))y_0\big);$$

the map is defined at each of the points $(-45, 2340)$, $(-45, -2340)$, $(1551, 59904)$, $(1551, -5990)$, and ∞_{C_1}:

$$\phi(-45, 2340) = \phi(1551, 59904) = \infty_{C_2}, \quad \phi(\infty_{C_1}) = (7, 133),$$
$$\phi(-45, -2340) = \left(-\tfrac{10898}{5187}, -\tfrac{477412081}{8968323}\right), \quad \phi(1551, -59904) = \left(\tfrac{16085}{5187}, \tfrac{477412081}{8968323}\right),$$

Here ∞_{C_1} is the point at infinity on C_1 and analogously for C_2. The map is regular at every point of C_1, so it is a morphism of curves. The inverse mapping is

$$\psi : C_1 \to C_2,$$

where $\psi(u, v) = (x_0, y_0)$ is given by

$$x_0 = \frac{5117 - 948u + 753u^2 + 266v}{(-7+u)^2},$$
$$y_0 = \frac{266\big(5201 + 9u(-4 + u(-22 + 13u))\big) + 2(1799 + 4797u)v}{(-7+u)^3},$$

which is not regular at the point ∞_{C_2} and is defined at the points $(7, 133)$ and $(7, -133)$:

$$\psi(7, 133) = \infty_{C_1}, \quad \psi(7, -133) = \left(-\tfrac{3628425}{17689}, \tfrac{8081948160}{2352637}\right).$$

With the notation $A = C_1 \backslash \{(-45, 2340), (1551, 59904)\}$, $B = C_2 \backslash \{\infty_{C_2}\}$, we have

$$\phi \circ \psi = \mathrm{id}_A, \quad \psi \circ \phi = \mathrm{id}_B .$$

We now give an explicit table of curves of rank at least 4 as stated in the Main Theorem. If we assume the parity conjecture we can show that some of them have actually the rank at least 5. Let

$$E_{(u^2-u-3)} : y^2 + (u^2 - u - 3)xy = x^3 + (u^2 - u - 3)x^2 - x + 1,$$

and $P_1 = (-309, 756)$, $P_2 = (-45, 2340)$, $T = (15, 0)$ — the points spanning the group $C_1(\mathbb{Q})$. From the computations above we can associate uniquely a pair (u, v) on C_2 corresponding to the point $\alpha T + \beta_1 P_1 + \beta_2 P_2$. We abbreviate this as $(u, v) \leftrightarrow (\alpha, \beta_1, \beta_2)$. We define the following functions:

- $R(u)$ is the regulator of the points

$$(0, 1), \ (1, 1), \ (u, u+1), \ \left(\tfrac{1}{9}, \tfrac{1}{54}(9 + 3u - 3u^2 + v)\right);$$

- $N(u)$ is the conductor of the curve $E_{(u^2-u-3)}$;
- $j(u)$ is the j-invariant of $E_{(u^2-u-3)}$;
- $w(u)$ is equal to the global root number $w(E_{(u^2-u-3)}/\mathbb{Q})$.

All the computations were performed for the minimal model of each curve.

For the last tuple the regulator is equal to 0 because the tuple corresponds to $u=\frac{1}{9}$ for which the fourth point from the statement of the Main Theorem coincides with the third point. For the tuple $(0,-1,1)$ (when $u=8/9$) the fourth point is linearly dependent on the other three points. Moreover the curves corresponding to these tuples are isomorphic over $\mathbb{Q}$.

Remark. We can find in the family $E_{(u^2-u-3)}$ curves of unconditional rank at least five. The curve $E_{239}: y^2+239xy=x^3+239x^2-x+1$ is a curve of unconditional rank five. The set of generators of the nontorsion part is given by

$$(0,1),\ (1,1),\ (16,17),\ \left(-\tfrac{14}{25},\tfrac{16661}{125}\right),\ \left(\tfrac{52}{81},\tfrac{469}{729}\right).$$

We can show that for $c=-\frac{14}{25}$ the associated auxiliary elliptic curve from the Main Theorem,

$$4-4c-3c^2+4c^3+2c^2u-c^2u^2-2c^2u^3+c^2u^4=w^2,$$

has rank 4 over $\mathbb{Q}$. Applying the technique of the proof of the Main Theorem we can actually prove a similar result to the one stated there. Precisely, we would have

α	β_1	β_2	$R(u)$	$N(u)$	$j(u)$	$W(u)$	rank
0	−2	−2	253637.08	7.42×10^{117}	−4382.17	−1	≥ 5
0	−2	−1	53400.57	4.79×10^{79}	-1.39×10^{6}	1	≥ 4
0	−2	0	16681.20	5.69×10^{59}	-4.14×10^{11}	−1	≥ 5
0	−2	1	23528.39	1.89×10^{64}	-1.11×10^{16}	1	≥ 4
0	−1	−2	117347.77	1.22×10^{95}	-4.66×10^{19}	1	≥ 4
0	−1	−1	6398.35	2.46×10^{46}	-7.42×10^{8}	1	≥ 4
0	−1	0	28.40	4.13×10^{12}	−1255.79	1	≥ 4
0	−1	1	0	6.54×10^{11}	−1264.95	1	–
0	0	−2	138113.04	6.46×10^{98}	−912.11	−1	≥ 5
0	0	−1	4697.68	1.21×10^{43}	−20742.18	1	≥ 4
0	0	0	8.61	57482738.0	-4.72×10^{9}	1	≥ 4
0	1	−2	608830.99	3.64×10^{145}	-4.45×10^{19}	1	≥ 4
0	1	−1	56796.71	1.80×10^{81}	-3.75×10^{10}	−1	≥ 5
0	1	0	1301.45	8.98×10^{31}	−200862.89	−1	≥ 5
0	1	1	0	6.54×10^{11}	−1264.95	1	–

Table 2. Curves of rank 4 and 5. See previous page for the meaning of the columns.

four linearly independent points for infinitely many rational parameters $u \in \mathbb{Q}$:

$$(0,1),\ (1,1),\ (u,u+1),\ \left(-\tfrac{14}{25},\ \tfrac{1}{125}(-105-35u+35u^2+v)\right),$$

where

$$v^2 = 17956+2450u-1225u^2-2450u^3+1225u^4.$$

Acknowledgments

The author thanks Wojciech Gajda for suggesting the problem; Sebastian Petersen, for helpful comments concerning root numbers and the parity conjecture; and Adam Lipowski, for the computational resources supporting Sage 3.4 [Stein et al. 2005]. The author also thanks the referee for several helpful suggestions on the mathematical content of the paper and the exposition.

References

[Brown and Myers 2002] E. Brown and B. T. Myers, "Elliptic curves from Mordell to Diophantus and back", *Amer. Math. Monthly* **109**:7 (2002), 639–649. MR 2003d:11080 Zbl 1083.11037

[Elkies 2007] N. D. Elkies, "Three lectures on elliptic surfaces and curves of high rank", preprint, 2007. arXiv 0709.2908

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. Erratum in **75** (1984), 381. MR 85g:11026a Zbl 0588.14026

[Kihara 1997] S. Kihara, "On an infinite family of elliptic curves with rank ≥ 14 over $\mathbf{Q}$", *Proc. Japan Acad. Ser. A Math. Sci.* **73**:2 (1997), 32. MR 98d:11059 Zbl 0906.11023

[Kowalski 2007] E. Kowalski, *Elliptic curves, rank in families and random matrices*, edited by J. B. Conrey et al., London Math. Soc. Lecture Note Series **341**, Cambridge University Press, 2007. MR 2008j:11073

[Mazur 1978] B. Mazur, "Rational isogenies of prime degree", *Invent. Math.* **44**:2 (1978), 129–162. MR 80h:14022 Zbl 0386.14009

[Mestre 1991] J.-F. Mestre, "Courbes elliptiques de rang ≥ 12 sur $\mathbf{Q}(t)$", *C. R. Acad. Sci. Paris Sér. I Math.* **313**:4 (1991), 171–174. MR 92m:11052 Zbl 0749.14026

[Nagao 1994] K.-i. Nagao, "An example of elliptic curve over $\mathbf{Q}(T)$ with rank ≥ 13", *Proc. Japan Acad. Ser. A Math. Sci.* **70**:5 (1994), 152–153. MR 95e:11064 Zbl 0848.14015

[Rohrlich 1994] D. E. Rohrlich, "Elliptic curves and the Weil–Deligne group", pp. 125–157 in *Elliptic curves and related topics*, edited by H. Kisilevsky and M. R. Murty, CRM Proc. Lecture Notes **4**, Amer. Math. Soc., Providence, RI, 1994. MR 95a:11054 Zbl 0852.14008

[Rubin and Silverberg 2007] K. Rubin and A. Silverberg, "Twists of elliptic curves of rank at least four", pp. 177–188 in *Ranks of elliptic curves and random matrix theory*, edited by J. B. Conrey et al., London Math. Soc. Lecture Note Ser. **341**, Cambridge Univ. Press, 2007. MR 2008e:11065 Zbl 05190710

[Shioda 1990] T. Shioda, "On the Mordell–Weil lattices", *Comment. Math. Univ. St. Paul.* **39**:2 (1990), 211–240. MR 91m:14056 Zbl 0725.14017

[Silverberg 2007] A. Silverberg, *The distribution of ranks in families of quadratic twists of elliptic curves*, edited by J. B. Conrey et al., London Math. Soc. Lecture Note Series **341**, Cambridge University Press, 2007. MR 2008c:11087

[Silverman 1983] J. H. Silverman, "Heights and the specialization map for families of abelian varieties", *J. Reine Angew. Math.* **342** (1983), 197–211. MR 84k:14033 Zbl 0505.14035

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Stein et al. 2005] W. Stein et al., Sage open-source mathematical software system, 2005, available at http://sagemath.org.

Received: 2010-01-30 Revised: 2010-08-25 Accepted: 2010-09-04

bartnas@amu.edu.pl *Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland*

INVOLVE 3:3(2010)

Curvature measures for nonlinear regression models using continuous designs with applications to optimal experimental design

Timothy O'Brien, Somsri Jamroenpinyo and Chinnaphong Bumrungsup

(Communicated by Kenneth S. Berenhaut)

We present and illustrate the methodology to calculate curvature measures for continuous designs, and extend design criteria to incorporate continuous designs. These design algorithms include quadratic design procedures, a subset design criterion, a second-order mean-square error design criterion, and a marginal curvature design methodology. A discussion of confidence intervals is also provided for continuous designs.

1. Introduction

When researchers have a given nonlinear regression model in mind to describe a specific process, they typically seek an experimental design to efficiently estimate the p model parameters with only negligible curvature. An additional requirement is usually that the design provide the researcher with the ability to test for the adequacy of the assumed model. Measures of curvature and nonlinearity are developed in [Beale 1960; Bates and Watts 1980; Ratkowsky 1983; Clarke 1987]; additional results and references are given in [Bates and Watts 1988; Seber and Wild 1989; Haines et al. 2004]. These measures assume that the chosen design has discrete weights, meaning that it is assumed that a sample size is fixed *a priori,* and that some fraction of these points is chosen at each of the design support points.

This paper argues that so-called continuous designs should instead be sought, and provides and demonstrates the means of calculating curvature measures for these designs. A by-product of this result is an extension to allow for continuous designs of the second-order mean-square error (MSE) design criterion in [Clarke and Haines 1995], the marginal curvature design strategy in [O'Brien 2006], and the second-order volume design criterion in [O'Brien 1992].

MSC2000: primary 53A45, 62J02; secondary 62K05.

Keywords: design measures, Fieller–Creasy problem, marginal curvature, model misspecification, nonlinearity, parameter subsets, robust designs.

Furthermore, most optimal design procedures provide designs with only p support points regardless of the sample size. Depending on the dispersion of the assumed prior distribution, one exception is the Bayesian procedures introduced in [Chaloner and Larntz 1989; Atkinson and Donev 1992]. OBrien [1992] shows how discrete $(p+1)$-point designs with reduced curvature could be obtained. In this article, we further extend this volume design strategy to incorporate continuous and larger discrete designs; these designs are recommended since they can also be used to test for lack-of-fit of the assumed model.

2. Notation and terminology

A general n-point design (measure) ξ can be written as

$$\xi = \left\{ \begin{array}{c} s_1, \ldots, s_r \\ \lambda_1, \ldots, \lambda_r \end{array} \right\}.$$

Here, the $r \leq n$ support points, $s_1, s_2, \ldots, s_r$ are elements of the design space $\boldsymbol{X}$, and the associated design weights — the λ's — are nonnegative real numbers which sum to one. When the support points are chosen from a discrete grid, we call this a discrete-point design; otherwise, it is a continuous-point design. Lattice and cyclic designs provide examples of the former type, whereas designs for dose-response curves or response surfaces are usually of the continuous variety. In addition, whenever the design weights are rational numbers proportional to $1/n$, we call the design a rational-mass (or rational-weight) design; otherwise, it is a continuous-mass design. Thus, even though ξ has n *design* points (which are not necessarily distinct), it may have only r distinct *support* points. Regardless of whether ξ is a continuous- or rational-mass design, $n\lambda_i$ is to be thought of as the *number of observations* taken at the experimental level s_i.

In regression settings, we advocate here that continuous point-continuous mass designs be obtained as a general rule-of-thumb, at least as a starting point; these designs can later be rounded to practical designs using the methodology given in [Pukelsheim and Rieder 1992]. For simplicity, we refer to these designs as *continuous designs* for the remainder of this paper. We encourage the use of continuous designs for two reasons. First, for some design criteria, optimality of the resulting design can be verified using a variation of the general equivalence theorem. This important result, first developed for linear models in [Kiefer and Wolfowitz 1960], and extended to nonlinear models in [White 1973], is illustrated for nonlinear models in [Haines 1992; O'Brien and Funk 2003]; important extensions to these original results are given in [Pukelsheim 1993; Atkinson and Haines 1996; Dette and O'Brien 1999]. This verification that the derived design is indeed optimal can be achieved by obtaining a graph of the corresponding variance function and noting whether or not this graph lies below a certain horizontal line. Since the

theorem is valid only for continuous designs (with continuous weights), verification of optimality of other designs is not possible in general.

Our second reason for preferring continuous designs is to expedite our search for an optimal design strategy in a given setting. For example, using the model function $\eta(x, \theta) = 1 - e^{-\theta x}$ and a specific prior distribution for θ, Atkinson [1988] provides discrete Bayesian D-optimal designs for sample sizes of $n = 1, 2$, and 3. The reported designs are $\{12\}$, $\{7, 27\}$, and $\{8, 8, 52\}$, respectively. Instead of continuing this trend of finding designs for larger sample sizes, we simply note that the continuous Bayesian D-optimal design associates the weight $\lambda_1 = 0.890$ with the point $s_1 = 9.05$ and the weight $\lambda_2 = 0.110$ with the point $s_2 = 120.1$. Thus, as we continue the sequence above up to the sample size $n = 9$, the first eight of these points would be placed nearer and nearer to $s_1 = 9.05$, whereas the last point approaches $s_2 = 120.1$. Thus, the strategy of first seeking continuous optimal designs gives us a clearer picture at the outset of what is required to achieve optimality.

3. Curvature measures for continuous designs

Under the usual assumption of uncorrelated Gaussian errors with zero mean and constant variance σ^2 (without loss of generality taken to equal one), the Fisher information *per observation* is given by

$$\boldsymbol{M}(\xi, \theta) = \sum_{k=1}^{r} \lambda_k \frac{\partial \eta(s_k)}{\partial \theta} \frac{\partial \eta(s_k)}{\partial \theta^T} = \boldsymbol{V}_{\boldsymbol{r}}^T \boldsymbol{\Lambda} \boldsymbol{V}_{\boldsymbol{r}},$$

where $\boldsymbol{V}_{\boldsymbol{r}}$ is the $r \times p$ Jacobian of η associated with ξ with kth row equal to $\partial \eta(s_k)/\partial \boldsymbol{\theta}^T$, and $\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, \ldots, \lambda_r\}$. It follows that the *total* information associated with an n-point design ξ is given by

$$n\boldsymbol{M}(\xi, \theta) = \sum_{k=1}^{r} n\lambda_k \frac{\partial \eta(s_k)}{\partial \theta} \frac{\partial \eta(s_k)}{\partial \theta^T} = \sum_{k=1}^{r} \frac{\partial \eta_w(s_k)}{\partial \theta} \frac{\partial \eta_w(s_k)}{\partial \theta^T} = \boldsymbol{V}_w^T \boldsymbol{V}_w.$$

In this expression,

$$\eta_w(s_k) = \sqrt{n\lambda_k}\,\eta(s_k) \tag{1}$$

is the weighted model function where the model weights correspond to the "number of observations" taken at the point s_k, and $\boldsymbol{V}_w$ is the $r \times p$ Jacobian of η_w with kth row equal to $\partial \eta_w(s_k)/\partial \boldsymbol{\theta}^T$. Thus, for continuous designs with n observations and r support points, and for second-order curvature, we point out that variance and volume measures should be based on

$$\boldsymbol{V}_w = \frac{\partial \eta_w(s, \theta)}{\partial \theta} = n^{1/2} \boldsymbol{\Lambda}^{1/2} \boldsymbol{V}_{\boldsymbol{r}}, \quad \boldsymbol{W}_w = \frac{\partial^2 \eta_w(s, \theta)}{\partial \theta \partial \theta^T} = n^{1/2} [\boldsymbol{\Lambda}^{1/2}][\boldsymbol{W}_{\boldsymbol{r}}]. \tag{2}$$

Here $\Lambda^{1/2} = \text{diag}\{\lambda_1^{1/2}, \ldots, \lambda_r^{1/2}\}$ and square bracket multiplication of arrays is defined in [Seber and Wild 1989, Appendix B]. The $r \times p$ matrix $\boldsymbol{V}_w$ plays the role of the Jacobian matrix and the $r \times p \times p$ array $\boldsymbol{W}_w$ that of the second derivative array; since the marginal curvature measures in [Clarke 1987] require third derivatives, the $r \times p \times p \times p$ array of third derivatives is defined analogously.

Example 1. Consider the Fieller–Creasy problem highlighted in [Cook and Witmer 1985; Clarke 1987; Seber and Wild 1989; Haines et al. 2004], in which the relevant nonlinear model function is $\eta(x, \theta_1, \theta_2) = \theta_1 x + \theta_1\theta_2(1 - x)$, with $x = 1$ for group A and $x = 0$ for group B. Interest lies in the parameter θ_2, which is the ratio of the group B mean over the group A mean. Here, we take $n_A = n\lambda$ and $n_B = n(1 - \lambda)$ for λ between 0 and 1; λ is thus the percentage of the experimental units assigned to group A. So as to adjust Wald confidence intervals to bring them more in line with likelihood intervals, [Clarke 1987] introduces marginal curvature measures Γ and β, where Γ assesses skewness of the interval and β (along with Γ) quantifies excessive kurtosis. Here, $\Gamma = \beta = 0$ is an indication of no nonlinearity, or exact coincidence between the two types of confidence intervals, as this is indeed the case for linear models. For the equal-weight case ($\lambda = 1/2$), these Γ and β expressions for θ_2 are given in [Haines et al. 2004, page 565]; extensions to the general case are given by the expressions

$$\Gamma = \frac{-2\theta_2\phi^{1/2}}{\theta_1\sqrt{n_A(1+\phi\theta_2^2)}} \quad \text{and} \quad \beta = \frac{1+2\phi\theta_2^2}{n_A\theta_1^2(1+\phi\theta_2^2)},$$

where $\phi = n_B/n_A = (1 - \lambda)/\lambda$. For example, when $n = n_A + n_B = 20$, $\theta_1 = 0.10$ and $\theta_2 = 1$, the equal-weight case results in $\Gamma = -4.4721$ and $\beta = 15.0$. In contrast, when $n_A = 19$ and $n_B = 1$ (so $\lambda = 0.95$), $\Gamma = -1.0260$ and $\beta = 5.0263$. This reduction in both Γ and β means that the Wald and likelihood intervals for θ_2 coincide to a greater degree in the latter situation than in the former (equal-weight) one.

Of course, both Γ and β for θ_1 are zero here since this parameter enters the model function a linear manner, so the two intervals will coincide for this parameter.

4. Applications to optimal design

An important benefit of extending the definitions of curvature measures to incorporate continuous designs is that more efficient designs can often then be obtained. Important examples include finding continuous designs for the Q-optimality, subset and MSE criteria, but first we examine discrete Q-optimal designs.

4.1. ***Discrete Q-optimal designs.*** For nonlinear models, we have from [Hamilton and Watts 1985] the second-order volume

$$v = c|\boldsymbol{V}_0^T\boldsymbol{V}_0|^{-1/2}|\boldsymbol{D}_0|^{-1/2}\{1 + k^2\,\text{tr}(\boldsymbol{D}_0^{-1}\boldsymbol{M}_0)\}, \tag{3}$$

as an approximation to the likelihood-based confidence region for the parameter vector $\boldsymbol{\theta}$. In this expression, c and k are constants with respect to the chosen design, matrices are evaluated at some initial guess $\boldsymbol{\theta}_0$ for $\boldsymbol{\theta}$. Also, $\boldsymbol{V}_0 = \boldsymbol{V}(\boldsymbol{\theta}_0)$ is the $n \times p$ Jacobian matrix whereas $\boldsymbol{M}_0$ and $\boldsymbol{D}_0 = \boldsymbol{I}_p - \boldsymbol{B}_0$ are associated with parameter-effects and intrinsic curvature, respectively. Whereas Hamilton and Watts [1985] obtain only designs with $n = p$, O'Brien [1992] shows how designs with $p + 1$ support points can be obtained to minimize the volume in (3), designs which are called discrete Q-optimal designs.

Specifically, the QR decomposition of $\boldsymbol{V}_0$ is

$$\boldsymbol{V}_0 = \boldsymbol{Q}_0\boldsymbol{R}_0 = [\boldsymbol{U}_0 \mid \boldsymbol{N}_0]\boldsymbol{R}_0 = \boldsymbol{U}_0\boldsymbol{L}_0^{-1},$$

so that the columns of $\boldsymbol{U}_0$ form an orthonormal basis for the tangent plane to $\eta(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and the columns of $\boldsymbol{N}_0$ form an orthonormal basis for the space orthogonal to this tangent plane. Since the residual vector is always orthogonal to the tangent plane at the maximum likelihood estimate, we can write

$$\boldsymbol{\varepsilon}_0 = \boldsymbol{N}_0\boldsymbol{\alpha}_0, \tag{4}$$

where $\boldsymbol{\alpha}_0$ is a vector of dimension $(n-p) \times 1$. Considering only the case $n = p+1$, O'Brien [1992] chooses the scalar α_0 equal to σ_0 since in general the expected squared length

$$E(\boldsymbol{\varepsilon}_0^T\boldsymbol{\varepsilon}_0) \approx E(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}) - E[(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T\boldsymbol{R}_0^T\boldsymbol{R}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] \approx n\sigma^2 - p\sigma^2. \tag{5}$$

We extend this result here for discrete designs with $n = p + s$ and $s \geq 2$.

Whenever $s = 2$ (i.e., $n = p + 2$), we can write $\boldsymbol{\alpha}_0 = \sqrt{2}\sigma_0\binom{\sin\phi}{\cos\phi}$, for some ϕ between 0 and $\pi/2$. With this choice for $\boldsymbol{\alpha}_0$, we keep its expected length identical to that of $\boldsymbol{\varepsilon}_0$ by (4) and (5) since $\boldsymbol{N}_0$ is an orthonormal matrix. With this choice of $\boldsymbol{\varepsilon}_0$, we can calculate the volume in (3) — subject to knowledge of ϕ. We thus define a $(p+2)$-point to be locally Q-optimal if it minimizes the expected volume

$$E[v(\phi)] = \int_0^{\pi/2} v(\phi)d\phi. \tag{6}$$

By extension, when $n = p + s$ for $s \geq 3$, we use hyperspherical coordinates for the s-vector $\boldsymbol{\alpha}_0$:

$$\alpha_0 = \sqrt{s}\sigma_0 \begin{pmatrix} \sin\phi_1 & \sin\phi_2 & \sin\phi_3 & \cdots & \sin\phi_{s-1} \\ \cos\phi_1 & \sin\phi_2 & \sin\phi_3 & \cdots & \sin\phi_{s-1} \\ & \cos\phi_2 & \sin\phi_3 & \cdots & \sin\phi_{s-1} \\ & & & & \vdots \\ & & & & \cos\phi_{s-1} \end{pmatrix}, \tag{7}$$

for $0 \le \varphi_1, \varphi_2, \ldots, \varphi_{s-1} \le \pi/2$. In this case, a discrete locally Q-optimal design minimizes the expected volume

$$\int_0^{\pi/2} \cdots \int_0^{\pi/2} \nu(\phi_1, \ldots, \phi_{s-1}) d\phi_1 \cdots d\phi_{s-1},$$

for $\nu = \nu(\varphi_1, \varphi_2, \ldots, \varphi_{s-1})$ given in (3), with $\boldsymbol{\varepsilon}_0$ in (4) and $\boldsymbol{\alpha}_0$ defined in (7). We illustrate this design strategy with the following example.

Example 2. To illustrate, consider the two-parameter intermediate product (IP2) model function

$$\eta(x, \boldsymbol{\theta}) = \frac{\theta_1}{\theta_1 - \theta_2} \{\exp(-\theta_2 x) - \exp(-\theta_1 x)\}, \tag{8}$$

for $\theta_1 > \theta_2 > 0$ and $x > 0$. This model function is used extensively in chemical kinetics and pharmacology. Using the initial parameter values $\boldsymbol{\theta}_0^T = (0.7, 0.2)$ and $\boldsymbol{\sigma}_0 = 0.1$, the 2-point discrete locally Q-optimal design takes one observation at each of $s_1 = 1.04$ and $s_2 = 5.59$, the 3-point discrete locally Q-optimal design takes one observation at each of $s_1 = 1.02$, $s_2 = 4.72$, and $s_3 = 6.81$, and the 4-point discrete locally Q-optimal design takes one observation at the points $s_1 = 1.00$, $s_2 = 1.23$, $s_3 = 5.35$, and $s_4 = 6.73$. This last design is obtained using the expected volume design expression given in (6). Table 1 contains the design points for the 2-point, 3-point and 4-point discrete locally Q-optimal designs for the intermediate product (IP2) function and the initial estimates $\boldsymbol{\theta}_0^T = (0.7, 0.2)$ and $\boldsymbol{\sigma}_0 = 0, 0.05, 0.10, 0.15$, and 0.20.

Not unexpectedly, the discrete Q-optimal designs for $\boldsymbol{\sigma}_0 = 0$, which coincide exactly with the local D-optimal designs, match exactly those given in [Box and Lucas 1959]. What may be surprising, though, is that for this example, whenever discrete Q-optimal designs were sought with five or more support points, the resulting designs had only four support points. Similar situations were observed

	2-pt design		3-pt design			4-pt design			
σ	s_1	s_2	s_1	s_2	s_3	s_1	s_2	s_3	s_4
0.00	1.23	6.86	1.23	6.86	6.86	1.23	1.23	6.86	6.86
0.05	1.15	6.36	1.15	5.79	7.11	1.12	1.26	6.18	6.97
0.10	1.04	5.59	1.02	4.72	6.81	1.00	1.23	5.35	6.72
0.15	0.94	4.88	0.91	4.24	6.24	0.90	1.18	4.74	6.34
0.20	0.81	4.05	0.82	3.92	5.52	0.82	1.13	4.34	5.91

Table 1. Design points associated with 2-, 3-, and 4-point discrete locally Q-optimal designs for the IP2 model function.

for other model functions. Thus, we conclude that in general, this discrete design strategy often produces designs with one or two extra design points.

4.2. ***Continuous Q-optimal designs.*** In (2), we showed how the first-order Jacobian matrix and second-order derivative array are extended to allow for continuous designs, and these are easily used in the calculation of the second-order volume approximation given in (3); designs which minimize this volume are called continuous Q-optimal designs.

Example 2 (continued). To illustrate, consider again the two-parameter intermediate product model given in expression (8) again with initial parameter estimates $\boldsymbol{\theta}_0^T = (0.7, 0.2)$ and $\boldsymbol{\sigma}_0 = 0.1$. In this case, the continuous locally Q-optimal design with $r = 3$ points associates the weights $\lambda = 0.46, 0.28, 0.26$ with the points $s = 1.06, 5.02, 6.73$, respectively. This latter design represents a 3.9% volume reduction in terms of (3) relative to the discrete 3-point Q-optimal design given in Section 4.1, and is therefore preferred. Table 2 contains the continuous locally Q-optimal designs with $n = 3$ for this model for $\boldsymbol{\sigma}_0 = 0, 0.05, 0.10, 0.15$, and 0.20. As pointed out above, techniques to obtain practical designs from design measures are discussed in [Pukelsheim and Rieder 1992].

In analogous manner to the performance of this model function with the discrete 5-point design criterion discussed above, continuous locally Q-optimal designs for $n > 3$ collapsed to those given in Table 2 for this model function.

Of course, this volume design strategy does not always yield designs with extra support points. One obvious counterexample is the Fieller–Creasy problem in which the dummy value takes on only two values — one for each of the two

σ	ξ		
0.00	$s = 1.23$	6.86	
	$\lambda = 0.50$	0.50	
0.05	$s = 1.17$	4.79	6.62
	$\lambda = 0.49$	0.04	0.47
0.10	$s = 1.06$	5.02	6.78
	$\lambda = 0.46$	0.28	0.26
0.15	$s = 0.96$	4.69	7.84
	$\lambda = 0.43$	0.47	0.10
0.20	$s = 0.82$	4.06	10.48
	$\lambda = 0.37$	0.58	0.05

Table 2. Continuous locally Q-optimal designs for the IP2 model function.

groups. Nonetheless, use of the criterion does reemphasize the importance of using continuous designs.

Example 1 (continued). For the Fieller-Creasy exercise, recall that λ is the proportion taken from group A, and the remaining $1-\lambda$ is the proportion taken from group B. It can be readily shown that since $n = p$ here, $\boldsymbol{D}_0$ is identity matrix, $\boldsymbol{I}_2$, the volume in (3) is proportional to

$$\lambda^{-1/2}(1-\lambda)^{-1/2}\{1+z/\lambda\}, \tag{9}$$

where

$$z = \frac{\sigma^2 \chi^2_{2,1-\alpha}}{4n\theta_1^2}.$$

Since the D-optimal design is obtained by taking $\sigma = 0$ in expression (9), this design is thus the equal-weight one; this design ignores curvature and minimizes only the generalized variance of $\boldsymbol{\theta}$. When $\sigma > 0$, the Q-optimal design is obtained by choosing

$$\lambda^* = \{1 - 4z + (16z^2 + 16z + 1)^{1/2}\}/4.$$

Contours of this optimal choice of the weight for group A are plotted in Figure 1 as a function of levels of noise $= \sigma^2/n$ and θ_1.

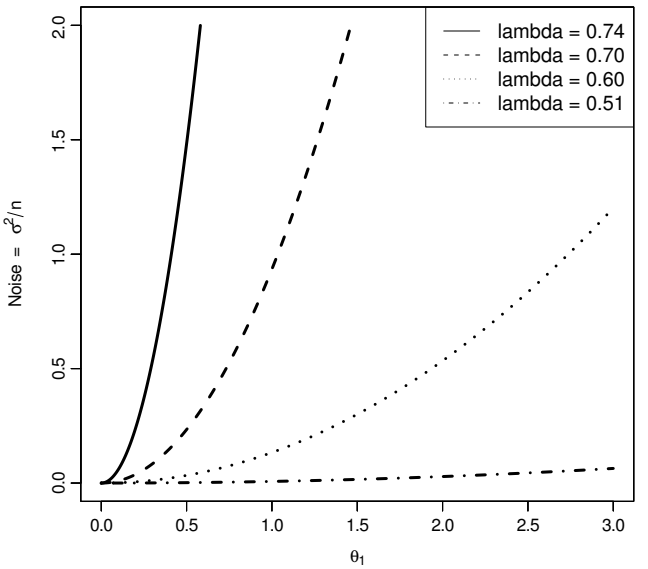


Figure 1. Contour plots of optimal lambda as a function of Noise $= \sigma^2/n$ and θ_1 for the Fieller–Creasy problem. Contour plots are of the form noise $= \boldsymbol{c}_\lambda \theta_1^2$, where $\boldsymbol{c}_\lambda$ is a constant depending upon λ. The horizontal line noise $= 0$ corresponds to $\lambda = 1/2$, whereas the vertical line $\theta_1 = 0$ corresponds to $\lambda = 3/4$.

In general, λ^* increases (from its lower bound of $1/2$) with the noise level, and its maximum value is $3/4$. This result is in line with the observation in Section 3 that the marginal curvature measures decrease as λ increases. Nevertheless, for values of λ above $3/4$, the generalized variance becomes exceedingly large.

4.3. ***Continuous optimal designs for parameter subsets.*** When the p-dimensional parameter vector is written $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ in situations where, without loss of generality, the p_1-vector $\boldsymbol{\theta}_1$ is the parameter (subset) vector of interest and the p_2-vector $\boldsymbol{\theta}_2$ are nuisance parameters, so that $p = p_1 + p_2$, then an optimal design can be sought using a subset design criteria. For example, Hill and Hunter [1974] have developed discrete subset D-optimal designs and given geometric interpretations of these designs. Atkinson and Donev [1992] extended this criterion to allow for continuous designs but did not give the geometric results; we provide them here.

Geometric aspects of continuous $D_s(\theta_1)$-optimal designs. We partition the $r \times p$ Jacobian $\boldsymbol{V_r}$ as

$$\boldsymbol{V_r} = [\boldsymbol{V}_1 \mid \boldsymbol{V}_2], \tag{10}$$

so $\boldsymbol{V}_1$ is $r \times p_1$ and corresponds to $\boldsymbol{\theta}_1$ and $\boldsymbol{V}_2$ is $r \times p_2$ and corresponds to $\boldsymbol{\theta}_2$. Then the Fisher information matrix can be written

$$\boldsymbol{M}(\xi, \theta) = \boldsymbol{V_r}^T \boldsymbol{\Lambda} \boldsymbol{V}_r = \begin{bmatrix} V_1^T \Lambda V_1 & V_1^T \Lambda V_2 \\ V_2^T \Lambda V_1 & V_2^T \Lambda V_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

and (local) $D_s(\boldsymbol{\theta}_1)$-optimal designs are those which maximize

$$|\boldsymbol{M}|/|\boldsymbol{M}_{22}| = |\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21}|. \tag{11}$$

provided $\boldsymbol{M}_{22}$ is nonsingular. In line with the above definitions, when the sample size n is set *a priori* and the design weights are scalar multiples of $1/n$, then the design is called discrete (mass); otherwise it is continuous (mass).

The case where $\boldsymbol{V}_1 = \boldsymbol{v}_1$ and $\boldsymbol{V}_2 = \boldsymbol{v}_2$ are column vectors (i.e., when $p_1 = p_2 = 1$) is of particular interest since it is easy to visualize the underlying objectives of the optimal design strategies. For example, Box and Lucas [1959] point out that in this instance the D-optimality criterion seeks designs which maximize the product of three terms: a term which captures the length of the vector $\boldsymbol{v}_1$, a term which captures the length of $\boldsymbol{v}_2$, and a term which increases with the angle between these two vectors. Hill and Hunter [1974] point out that the subset criterion in (11) chooses designs to maximize only the product of the first and third of these terms.

Example 2 (continued). For the IP2 model setting described above, note that the discrete $D_s(\theta_1)$-optimal design associates that weight $\lambda = 1/2$ with each of the points $s_1 = 1.172$ and $s_2 = 7.441$, and the column vectors are plotted in Figure 2 and denoted $\boldsymbol{v}_1^D$ and $\boldsymbol{v}_2^D$ respectively.

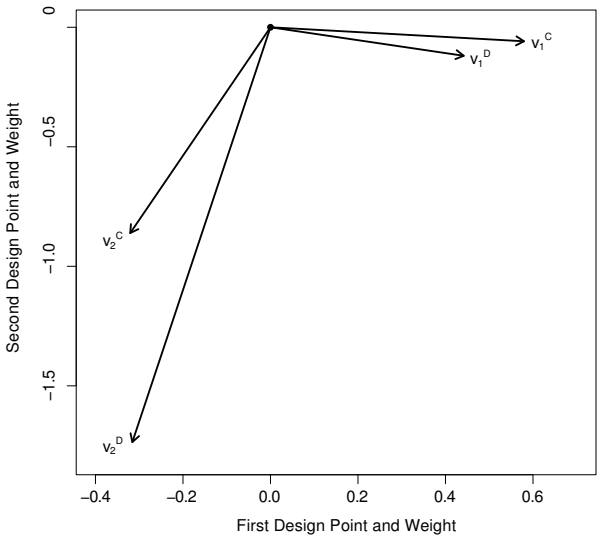


Figure 2. IP2 model and visualization of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ for the $D_s(\theta_1)$-optimal (subset) continuous (C) and discrete (D) designs.

If we take $n = 2$, this is the same result we obtain if we use the $\boldsymbol{V}_W$ matrix in (2): precisely what is used to derive analogous results for the continuous subset design. This latter design associates the weight $\lambda_1 = 0.878$ with the point $s_1 = 0.994$, and the weight $\lambda_2 = 0.122$ with the point $s_2 = 7.122$; the corresponding vectors, denoted $\boldsymbol{v}_1^C$ and $\boldsymbol{v}_2^C$, are also plotted in Figure 2. In terms of the measure in (11), the efficiency of the discrete design with respect to the continuous one is only 0.6549 — that is, about 65%. Since $1/0.6549 = 1.53$, this means that the discrete design needs to be 53% larger than the continuous one in order to yield the same amount of information as the continuous design. Geometrically, as we shift from the discrete design ($\boldsymbol{v}_1^D$ and $\boldsymbol{v}_2^D$) to the continuous one ($\boldsymbol{v}_1^C$ and $\boldsymbol{v}_2^C$) in Figure 2, the small reduction in orthogonality between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in the figure results in a significant lengthening of the $\boldsymbol{v}_1$ vector.

Before we leave this example, let's now extend the concept of *design locus* presented in [Hill and Hunter 1974, page 430] and [Atkinson and Donev 1992, page 200] to allow for continuous designs. Whereas the geometric strategy in Figure 2 is to focus on the columns of $\boldsymbol{V}_W$, the design loci, graphed in Figure 3, focus on the rows of this matrix.

For this example, the solid curve in the center of Figure 3 corresponds to the discrete design and is traced out as x ranges from 0 to infinity; the two solid circles on this design locus correspond to the discrete design support points given above. In contrast, the outer dot-dashed curve is the design locus for $\lambda_1 = 0.878$ and the inner dotted curve for $\lambda_2 = 0.122$, with the corresponding continuous design points

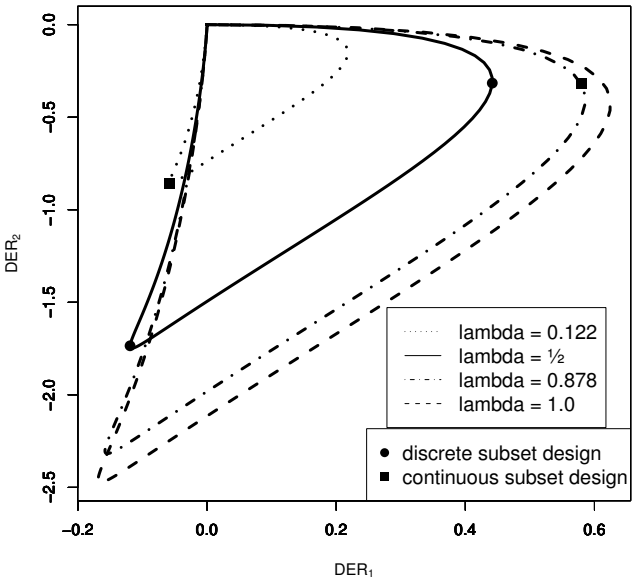


Figure 3. Design loci plots of $\mathrm{DER}_2 = \partial\eta_w/\partial\theta_2 = \sqrt{\eta\lambda}\,\partial\eta/\partial\theta_2$ versus $\mathrm{DER}_1 = \partial\eta_w/\partial\theta_1 = \sqrt{\eta\lambda}\,\partial\eta/\partial\theta_1$ for $\lambda = 0.122, 1/2, 0.878, 1.0$. Points correspond to the discrete subset design (filled with circles on the $\lambda = 1/2$ design locus) and the continuous subset design (filled with squares on the $\lambda = 0.122$ and 0.878 design loci).

indicated on the respective loci (darkened squares). For completeness, the design locus corresponding to $\lambda = 1$ — which represents the outermost possible design locus — is plotted in the figure with the dashed curve.

Continuous $D_2(\theta_1)$-optimal designs with reduced curvature. The results presented in Sections 4.1 and 4.2 underscore the importance of seeking optimal designs for nonlinear models with reduced curvature in addition to smaller variance, and we now consider this issue in the context of subset designs. As mentioned above, three design criteria which take account of nonlinearity are the volume criterion of [Hamilton and Watts 1985; O'Brien 1992], the second-order MSE criterion of [Clarke and Haines 1995], and the marginal curvature criterion of [O'Brien 2006]. The last of these criteria is well suited for subset designs and interested readers are referred to see [O'Brien 2006]; an extension of the volume criterion to subset designs is not as straightforward, and is not considered here. Rather, we underscore here the use of continuous designs with the MSE criterion.

Employing the second-order variance approximation given in [Clarke 1980] and the second-order bias approximation given in [Box 1971], Clarke and Haines [1995] illustrate the use of discrete subset second-order minimum MSE designs.

We call these designs (local) $D_2(\theta_1)$-optimal here. This criterion is well suited to incorporating continuous design as is demonstrated in the following illustration.

Example 2 (continued). For the IP2 model setting described above, discrete and continuous $D_s(\theta_1)$-optimal designs are given in the previous section, however we point out that these designs do not attempt to reduce nonlinearity. In contrast, the continuous $D_2(\theta_1)$-optimal design, which associates the weight $\lambda_1 = 0.825$ with the point $s_1 = 0.834$ and the weight $\lambda_2 = 0.175$ with the point $s_2 = 5.706$, does result in lower curvature measures. We return to the assessment of curvature in situations such as the present one in the Discussion. Interestingly, in comparing this continuous subset design with the first-order ($D_s(\theta_1)$-optimal) one, although the design weights remain essentially unchanged, the design support points shift downward towards the origin.

5. Applications to confidence intervals and regions

Important confidence regions for model parameters associated with nonlinear models using discrete designs are the Wald, likelihood-based and second-order regions of the form $\{\boldsymbol{\theta}$ such that LHS $\leq \rho_a^2\}$. The left hand sides (LHS's) in this expression are as follows:

$$\begin{aligned} \text{Wald : LHS} &= (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^T \boldsymbol{V}_*^T \boldsymbol{V}_* (\boldsymbol{\theta} - \boldsymbol{\theta}_*), \\ \text{Likelihood : LHS} &= S(\boldsymbol{\theta}) - S(\boldsymbol{\theta}_*), \\ \text{Second-order : LHS} &= \boldsymbol{\tau}^T (\boldsymbol{I}_p - \boldsymbol{B}) \boldsymbol{\tau}. \end{aligned} \tag{12}$$

The first two of these regions are discussed in [Seber and Wild 1989]. The last of these regions, developed in [Hamilton et al. 1982], leads directly to the second-order volume design criterion given in [Hamilton and Watts 1985] since the given volume approximates the volume of this second-order region. In these expressions, $\boldsymbol{\theta}_*$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, $S(\boldsymbol{\theta}) = \Sigma[y_k - \eta_k(\boldsymbol{\theta})]^2$ is the usual sum-of-squares function, $\boldsymbol{B}$ is as in (3), and

$$\boldsymbol{\tau} = \boldsymbol{U}^T [\eta(\boldsymbol{\theta}) - \eta(\boldsymbol{\theta}_*)]$$

for $\boldsymbol{U}$ discussed in the paragraph following (3). Also, based on [Seber and Wild 1989, page 261], we take $\rho_a^2 = \sigma^2 \chi_p^2$ since σ is assumed to be known. With the definitions given in Section 3 we now extend these regions to allow for continuous designs.

Specifically, we use the weighted Jacobian matrix given in (2) for the Wald region, the weighted model function in (1) in the calculations of the sum-of-squares functions used in the likelihood-based region, and the weighted Jacobian matrix and Hessian array in (2) in the calculations for the second-order region. The following example provides an illustration.

Example 3. Using the one-parameter simple exponential (SE1) model function, $\eta(x,\theta) = e^{-\theta x}$, and initial parameter estimates $\theta_0 = 0.10$ and $\sigma_0 = 0.40$, the two-point discrete Q-optimal design associates the weight $\lambda = 1/2$ with each of the points $s = 5.10$ and 7.34, whereas the continuous Q-optimal design associates the weight $\lambda_1 = 0.926$ with the point $s_1 = 5.49$ and the weight $\lambda_2 = 0.074$ with the point $s_2 = 25.92$. Then, for this continuous design, the left-hand sides (LHS) in (12), that is, the so-called confidence functions, are given as follows:

$$\begin{aligned}
\text{Wald : LHS} &= 19.163(\theta - 0.10)^2, \\
\text{Likelihood : LHS} &= (0.85 - 1.36e^{-5.49\theta})^2 + (-0.37 - 0.39e^{-25.9\theta})^2 - 0.40^2, \\
\text{Second-order : LHS} &= 1.314(1.341e^{-5.49\theta} + 0.066e^{-25.9\theta} - 0.780)^2.
\end{aligned}$$

These confidence functions are plotted for this continuous design in the left panel of Figure 4 along with the horizontal cut-line $\rho_a^2 = \sigma^2\chi^2_{1,0.90} = 0.40^2 2.71 = 0.434$.

The corresponding confidence intervals are thus the intervals on the cut-line between the sides of the respective confidence functions. The inadequacy of the

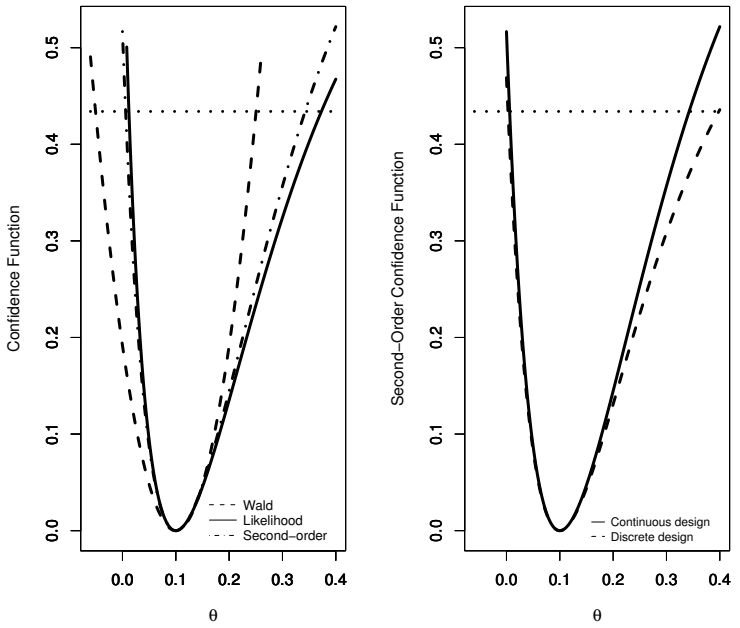


Figure 4. Confidence functions for the SE1 (one-parameter simple exponential) model function. In the left panel, plot of the three confidence interval methods in (12). In the right panel, plot of the second-order (Hamilton and Watts) function for the continuous and discrete Q-optimal designs. Horizontal cut-line corresponds to 90% confidence.

Wald interval is readily apparent here. Also, note the considerable coincidence here between the second-order and likelihood functions (and thus confidence intervals).

In the right panel of Figure 4, the second-order confidence functions are plotted for the continuous and discrete Q-optimal designs. That the confidence interval for the continuous design is shorter (by about 15%) than the confidence interval for the discrete design attests to the efficiency of the former design and underscores the importance of using continuous designs.

6. Discussion

This article demonstrates how simple yet important adjustments to the model function in (1) and to the associated Jacobian and other matrices given in (2) can be made so as to permit the use of continuous designs in curvature measures, optimal design procedures and the calculation of confidence intervals. The gains of continuous designs over discrete designs, wherein the final sample size is assumed fixed *a priori*, are often significant. As a result, we recommend that continuous designs always be obtained and examined at least as a starting point.

As mentioned above in Section 4, several design criteria exist which can be used to obtain designs with reduced curvature. The Q-optimality criterion, used extensively in this paper, provides designs that minimize the second-order approximation to the volume of the likelihood confidence region. The MSE criterion of [Clarke and Haines 1995] provides designs that minimize a second-order approximation of the mean-square error of the least-squares parameter estimate. The MC (marginal curvature) approach of [O'Brien 2006] yields designs which minimize the first-order variance approximation and the marginal curvature of [Clarke 1987] for a specified parameter or set of parameters. The latter two of these criteria are readily adapted to yield efficient designs for parameter subsets, as demonstrated on p. 327. Clearly more work needs to be performed exploring which of these three criteria is "best". It may well be the case, however, that each criterion does well only for the respective curvature measure (but not the others), and how one determines which is best may have more to do with how one defines curvature or nonlinearity.

A clear advantage of the volume criterion developed in [O'Brien 1992] and extended here is that this design strategy usually yields designs with extra support points which can be used to test for model misspecification. The same beneficial result is associated with the design strategy in [O'Brien 2006], which focuses on geometric designs. Govaerts [1996] correctly calls an optimal design's having only p support points a "big limitation" in practical settings. This is especially important when one keeps in mind George Box's comment that "all models are wrong but some are useful" [Box 1979]. These arguments clearly lead us to prefer the Q-optimality and MC design approaches discussed here.

References

[Atkinson 1988] A. C. Atkinson, "Recent developments in the methods of optimum and related experimental designs", *Internat. Statist. Rev.* **56**:2 (1988), 99–115. MR 89g:62123 Zbl 0646.62064

[Atkinson and Donev 1992] A. C. Atkinson and A. N. Donev, *Optimum experimental designs*, Oxford Statistical Science Series **8**, Clarendon Press, 1992. Zbl 0829.62070

[Atkinson and Haines 1996] A. C. Atkinson and L. M. Haines, "Designs for nonlinear and generalized linear models", pp. 437–475 in *Design and analysis of experiments*, edited by R. Ghosh and C. Rao, Handbook of Statist. **13**, North-Holland, Amsterdam, 1996. MR 1492576 Zbl 0910.62070

[Bates and Watts 1980] D. M. Bates and D. G. Watts, "Relative curvature measures of nonlinearity", *J. Roy. Statist. Soc. Ser. B* **42**:1 (1980), 1–25. MR 81e:62070 Zbl 0455.62028

[Bates and Watts 1988] D. M. Bates and D. G. Watts, *Nonlinear regression analysis and its applications*, Wiley, New York, 1988. MR 92f:62002 Zbl 0728.62062

[Beale 1960] E. M. L. Beale, "Confidence regions in non-linear estimation", *J. Roy. Statist. Soc. Ser. B* **22** (1960), 41–88. MR 22 #10055 Zbl 0096.13201

[Box 1971] M. J. Box, "Bias in nonlinear estimation", *J. Roy. Statist. Soc. Ser. B* **33** (1971), 171–201. MR 47 #4376 Zbl 0232.62029

[Box 1979] G. E. P. Box, "Robustness in the strategy of scientific model building", pp. 201–236 in *Robustness in statistics* (Triangle Park, NC, 1978), edited by R. L. Launer and G. N. Wilkinson, Academic Press, New York, 1979.

[Box and Lucas 1959] G. E. P. Box and H. L. Lucas, "Design of experiments in non-linear situations", *Biometrika* **46**:1/2 (1959), 77–90. MR 21 #949 Zbl 0086.34803

[Chaloner and Larntz 1989] K. Chaloner and K. Larntz, "Optimal Bayesian design applied to logistic regression experiments", *J. Statist. Plann. Inference* **21**:2 (1989), 191–208. MR 90f:62241 Zbl 0666.62073

[Clarke 1980] G. P. Y. Clarke, "Moments of the least squares estimators in a nonlinear regression model", *J. Roy. Statist. Soc. Ser. B* **42**:2 (1980), 227–237. MR 81j:62117

[Clarke 1987] G. P. Y. Clarke, "Marginal curvatures and their usefulness in the analysis of nonlinear regression models", *J. Amer. Statist. Assoc.* **82**:399 (1987), 844–850. MR 88k:62104 Zbl 0623.62062

[Clarke and Haines 1995] G. P. Y. Clarke and L. M. Haines, "Optimal design for models incorporating the Richards function", pp. 61–66 in *Statistical modelling* (Innsbruck, 1995), edited by G. U. H. Seeber et al., Lecture Notes in Statistics **104**, Springer, New York, 1995.

[Cook and Witmer 1985] R. D. Cook and J. A. Witmer, "A note on parameter-effects curvature", *J. Amer. Statist. Assoc.* **80**:392 (1985), 872–878. MR 819586

[Dette and O'Brien 1999] H. Dette and T. E. O'Brien, "Optimality criteria for regression models based on predicted variance", *Biometrika* **86**:1 (1999), 93–106. MR 2000a:62171 Zbl 1101.62357

[Govaerts 1996] B. Govaerts, "Discussion of the papers by Atkinson, and Bates *et al*", *J. Royal Statistical Soc. B* **58**:1 (1996), 104.

[Haines 1992] L. M. Haines, "Optimal design for inverse quadratic polynomials", *South African Statist. J.* **26**:1 (1992), 25–41. MR 93j:62189 Zbl 0763.62042

[Haines et al. 2004] L. M. Haines, T. E. O'Brien, and G. P. Y. Clarke, "Kurtosis and curvature measures for nonlinear regression models", *Statist. Sinica* **14**:2 (2004), 547–570. MR 2059296 Zbl 1045.62065

[Hamilton and Watts 1985] D. C. Hamilton and D. G. Watts, "A quadratic design criterion for precise estimation in nonlinear regression models", *Technometrics* **27**:3 (1985), 241–250. MR 797562 Zbl 0598.62083

[Hamilton et al. 1982] D. C. Hamilton, D. G. Watts, and D. M. Bates, "Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions", *Ann. Statist.* **10**:2 (1982), 386–393. MR 84i:62082 Zbl 0537.62045

[Hill and Hunter 1974] W. J. Hill and W. G. Hunter, "Design of experiments for subsets of parameters", *Technometrics* **16** (1974), 425–434. MR 50 #11655 Zbl 0311.62045

[Kiefer and Wolfowitz 1960] J. Kiefer and J. Wolfowitz, "The equivalence of two extremum problems", *Canad. J. Math.* **12** (1960), 363–366. MR 22 #8616 Zbl 0093.15602

[O'Brien 1992] T. E. O'Brien, "A note on quadratic designs for nonlinear regression models", *Biometrika* **79** (1992), 847–849.

[O'Brien 2006] T. E. O'Brien, "Robust optimal design with reduced curvature", in *Proceedings of the American Statistical Association, Biopharmaceutical Section* (Alexandria, VA), 2006.

[O'Brien and Funk 2003] T. E. O'Brien and G. M. Funk, "A gentle introduction to optimal design for regression models", *Amer. Statist.* **57**:4 (2003), 265–267. MR 2037854

[Pukelsheim 1993] F. Pukelsheim, *Optimal design of experiments*, Wiley, New York, 1993. MR 94k:62124 Zbl 0834.62068

[Pukelsheim and Rieder 1992] F. Pukelsheim and S. Rieder, "Efficient rounding of approximate designs", *Biometrika* **79**:4 (1992), 763–770. MR 1209476

[Ratkowsky 1983] D. A. Ratkowsky, *Nonlinear regression modeling: A unified practical approach*, Marcel Dekker, New York, 1983. Zbl 0572.62054

[Seber and Wild 1989] G. A. F. Seber and C. J. Wild, *Nonlinear regression*, Wiley, New York, 1989. MR 90j:62004 Zbl 0721.62062

[White 1973] L. V. White, "An extension of the general equivalence theorem to nonlinear models", *Biometrika* **60** (1973), 345–348. MR 48 #5290 Zbl 0262.62037

Received: 2010-05-19 Accepted: 2010-06-13

tobrien@math.luc.edu *Department of Mathemtaics and Statistics, Loyola University, 1032 W. Sheridan Road, Chicago, IL 60660-1537, United States*
http://webpages.math.luc.edu/~tobrien/home.html

huajam@grad.sci.tu.ac.th *Department of Mathematics and Statistics, Thammasat University — Rangsit Campus, Khlongluang, Pathumthani 12121, Thailand*

chinnaphong@mathstat.sci.tu.ac.th *Department of Mathematics and Statistics, Thammasat University — Rangsit Campus, Khlongluang, Pathumthani 12121, Thailand*

INVOLVE 3:3(2010)

Numerical semigroups from open intervals

Vadim Ponomarenko and Ryan Rosenbaum

(Communicated by Scott Chapman)

We consider numerical semigroups $\mathbb{N} \cap \mathbb{N}I$, for intervals I. We compute the Frobenius number and multiplicity of such semigroups, and show that we may freely restrict I to be open, closed, or half-open, as we prefer.

Given an interval $I \subseteq \mathbb{Q}^{>0}$ in the positive rationals, consider the set

$$S(I) = \mathbb{N} \cap \mathbb{N}I = \{m \in \mathbb{N} : \exists\, n \in \mathbb{N}, m/n \in I\}.$$

This turns out to be a numerical semigroup, and has been the subject of considerable recent investigation (see [Rosales and García-Sánchez 2009, Chapter 4] for an introduction). Special cases include modular numerical semigroups [Rosales et al. 2005], where $I = [m/n, m/(n-1)]$ $(m, n \in \mathbb{N})$; proportionally modular numerical semigroups [Rosales et al. 2003], where $I = [m/n, m/(n-s)]$ $(m, n, s \in \mathbb{N})$; and opened modular numerical semigroups [Rosales and Urbano-Blanco 2006] where $I = (m/n, m/(n-1))$ $(m, n \in \mathbb{N})$.

We consider instead arbitrary open intervals $I = (a, b)$. We show that this set of semigroups coincides with the set of semigroups generated by closed and half-open intervals. Consequently, this class of semigroups contains modular numerical semigroups, proportionally modular numerical semigroups, as well as opened modular numerical semigroups. We also compute two important invariants of these numerical semigroups: the Frobenius number $g(S(I))$ and multiplicity $m(S(I))$.

1. Preliminaries

We begin by defining a helpful function $\kappa(a, b)$. For $a, b \in \mathbb{R}$ with $a < b$ we define $\kappa(a, b) = \lfloor b/(b-a) \rfloor$. The function κ has various nice properties, for example $\kappa(a, b) = \kappa(ac, bc)$ for $c > 0$. In the special case of $a = m/n$, $b = m/(n-s)$, we have $\kappa(a, b) = \lfloor n/s \rfloor$. The following properties of $\kappa(a, b)$ are needed in the sequel.

Lemma 1.1. *Let $a, b \in \mathbb{R}$ with $a < b$ and $b \neq 0$. Set $\kappa = \kappa(a, b)$. If $\kappa \neq 0$, then $(\kappa - 1)/\kappa \leq a/b$. If $\kappa \neq -1$, then $a/b < \kappa/(\kappa + 1)$.*

MSC2000: 20M10, 20M14.

Keywords: numerical semigroup, modular numerical semigroup.

Proof. We have $\kappa \le b/(b-a) < \kappa+1$. Assume that $\kappa \notin \{-1, 0\}$. Then

$$\kappa, \frac{b}{b-a}, \kappa+1$$

all have the same sign, and we have $1/\kappa \ge (b-a)/b > 1/(\kappa+1)$, hence

$$\frac{1-\kappa}{\kappa} \ge -\frac{a}{b} > \frac{-\kappa}{\kappa+1},$$

and the results follow. If $\kappa = 0$, then $0 \le b/(b-a) < 1$, so $b > 0$ and $b < b-a$ so $a < 0$ hence $a/b < 0 = \kappa/(\kappa+1)$. If $\kappa = -1$, then $-1 \le b/(b-a) < 0$, so $b < 0$ and $a - b \le b$ so $a \le 2b$ and $a/b \ge 2 = (\kappa-1)/\kappa$. □

Lemma 1.2. *Let $a, b \in \mathbb{R}$ with $a < b$ and $b > 0$. Then*

$$\mathbb{N} \setminus S((a,b)) = \mathbb{N} \cap \bigcup_{n=1}^{\kappa(a,b)} [b(n-1), an].$$

Proof. Because $S((a,b)) = \mathbb{N} \cap \cup_{n=1}^{\infty}(an, bn)$, we have

$$\mathbb{N} \setminus S((a,b)) = \mathbb{N} \cap \bigcup_{n=1}^{\infty} [b(n-1), an].$$

Since $b > 0$, $\kappa(a,b) \ne -1$ and hence by Lemma 1.1, $b\kappa(a,b) > a(\kappa(a,b)+1)$. Hence for $n > \kappa(a,b)$, the intervals are empty and may be excluded. □

Lemma 1.2 yields an upper bound for g. This bound will later be improved in Theorem 3.1, but for the purposes of Theorem 2.3 the following weaker bound suffices.

Corollary 1.3. *Suppose $0 < a < b$. Then $g(S((a,b))) \le \lfloor a\kappa(a,b) \rfloor$.*

2. Intervals

We now prove that restricting I to be open is harmless, as this class of semigroups coincides with ones generated by closed or half-open intervals. To reduce the number of cases to consider, we introduce the symbols $\{, \}$ to denote endpoints of an interval that are either open or closed. For example, $(a, b\}$ indicates an interval that is open on the left. The meaning of these symbols is determined when first used, and then remains consistent; that is, if $(a, b\}$ is open then $[a', b\}$ means $[a', b)$, and if $(a, b\}$ is half-open then $[a', b\}$ means $[a', b]$.

The following lemma is the cornerstone of the interval equivalence results. Let $d(x)$ denote the denominator of reduced rational x.

Lemma 2.1. *Let $a \in \mathbb{Q}^{>0}$, $n \in \mathbb{N}$. Then all rationals in the interval*

$$\Big(a - \frac{a}{nd(a)+1}, a + \frac{a}{nd(a)+1}\Big),$$

other than possibly a, have numerator greater than n.

Proof. Suppose $a = p/q$, so $d(a) = q$. Consider any rational x/y with

$$0 < \left|\frac{x}{y} - a\right| < \frac{a}{nq+1} = \frac{p}{q(nq+1)}.$$

Also, we have

$$\left|\frac{x}{y} - \frac{p}{q}\right| = \left|\frac{xq - yp}{yq}\right| \geq \frac{1}{yq},$$

because $xq - yp \neq 0$ since $x/y \neq a$. Combining, we get

$$\frac{p}{q(nq+1)} > \frac{1}{yq},$$

hence

$$\frac{px}{nq+1} > \frac{x}{y} > a - \frac{a}{nq+1} = \frac{pnq}{q(nq+1)} = \frac{pn}{nq+1},$$

and thus $x > n$. □

Lemma 2.2. *Suppose that I, J, $I \cup J$ are all intervals. Then $S(I \cup J) = S(I) \cup S(J)$. Also, if $I \subseteq J$, then $S(I) \subseteq S(J)$ and $g(S(I)) \geq g(S(J))$.*

Proof. An integer $m \in S(I \cup J)$ if and only if $m/n \in I \cup J$ for some n. This is true if and only if $m/n \in I$ or $m/n \in J$. Hence $m \in S(I \cup J)$ if and only if $m \in S(I) \cup S(J)$. If $I \subseteq J$, then $S(J) = S(I \cup J) = S(I) \cup S(J) \supseteq S(I)$. □

The following theorem lets us replace a closed endpoint with an open one nearby, leaving the semigroup unchanged. Given a modular or proportionally modular numerical semigroup S, it explicitly gives an open interval I with $S(I) = S$.

Theorem 2.3. *Let $0 < a < b$. Set*

$$a' = \begin{cases} a - \dfrac{a}{\lfloor a\kappa(a,b)\rfloor d(a)+1} & \text{if } a \in \mathbb{Q}, \\ a & \text{if } a \notin \mathbb{Q}, \end{cases}$$

$$b' = \begin{cases} b + \dfrac{b}{\lfloor a\kappa(a,b)\rfloor d(b)+1} & \text{if } b \in \mathbb{Q}, \\ b & \text{if } b \notin \mathbb{Q}. \end{cases}$$

Then

$$S([a,b\}) = S((a',b\}) \quad \text{and} \quad S(\{a,b]) = S(\{a,b']).$$

Proof. We consider only $[a,b\}$; $\{a,b]$ is symmetric. Suppose first that $a \notin \mathbb{Q}$. By Lemma 2.2, $S([a',b\}) = S((a',b\}) \cup S([a',a']) = S((a',b\})$ since $S([a',a']) = \varnothing$. We now assume $a \in \mathbb{Q}$. Since $a < 2a - a'$, Lemma 2.2 implies that $S((a',b\}) = S((a',2a-a')) \cup S([a,b\})$. We will show $S((a',2a-a')) \subseteq S([a,b\})$, implying $S((a',b\}) \subseteq S([a,b\})$ (and $S((a',b\}) \supseteq S([a,b\})$ by Lemma 2.2).

Let $c \in S((a',2a-a'))$. Hence there is some $d \in \mathbb{N}$ so that $c/d \in (a',2a-a')$. By Lemma 2.1, either $c/d = a$ (in which case $c \in S([a,b\})$), or $c > \lfloor a\kappa(a,b)\rfloor$.

In the latter case, we apply Corollary 1.3 and $c > \lfloor a\kappa(a,b)\rfloor \geq g(S((a,b))) \geq g(S([a,b\})))$, so $c \in S([a,b\})$. □

Theorem 2.4 is a counterpoint to Theorem 2.3, allowing us to replace an open endpoint with a closed one nearby. Proposition 5 in [Rosales and Urbano-Blanco 2006] tells us more: that every $S(I)$ is proportionally modular; that is, there are $m, n, s \in \mathbb{N}$ where $S(I) = S([m/n, m/(n-s)])$. Unfortunately neither of these results give an explicit formula such as in Theorem 2.3.

Theorem 2.4. *Let $0 < a < b$. Then there are a', b' with $S((a,b\}) = S([a',b\})$ and $S(\{a,b)) = S(\{a',b]) $. Further, $a/a', b/b' \in \mathbb{Q}$.*

Proof. We consider only $(a,b\}$; $\{a,b)$ is symmetric. Suppose first that $a \notin \mathbb{Q}$. By Lemma 2.2, $S([a',b\}) = S((a',b\}) \cup S([a',a']) = S((a',b\})$ since $S([a',a']) = \varnothing$. We now assume $a \in \mathbb{Q}$. Let a_0 be any rational in (a,b), and consider the sequence given by $a_i = \frac{1}{2}(a + a_{i-1})$, for $i \geq 1$. By Lemma 2.2, we have $S([a_1,b\}) \subseteq S([a_2,b\}) \subseteq \cdots \subseteq S((a,b\})$. Set

$$X = S((a,b\}) \setminus S([a_1,b\}),$$

a finite set. Set $Z = \{x/y : x \in X, x/y \in (a,b\}\}$. Since $a_i \to a$ and $\min Z > a$, there is some $j > 0$ with $Z \subseteq [a_j,b\}$, and hence $X \subseteq S([a_j,b\})$. We take $a' = a_j$; note that $a' \in \mathbb{Q}$ by construction. □

3. Calculating $g(S((a,b)))$ and $m(S((a,b)))$

We now improve on Corollary 1.3 by calculating $g(S((a,b)))$ exactly. Various other results are known in related contexts. For example, if $S([a,b])$ is not a half-line, in [Rosales and Urbano-Blanco 2006] it was shown that

$$\frac{g(S([a,b]))}{g(S([a,b])) - 1} < a < b < g(S([a,b])).$$

Also, if $2 \leq a < b$ with $a, b \in \mathbb{N}$, in [Rosales and Vasco 2009] it was shown that $g(S((a,b))) = b$.

Theorem 3.1. *Suppose $0 < a < b$. Set $\kappa = \kappa(a,b)$, $\kappa' = max(\kappa(a-1,b-1),0)$. Then $g(S((a,b))) = \lfloor a\alpha\rfloor$, where $\alpha \in \mathbb{Z}$ satisfies $\kappa' \leq \alpha \leq \kappa$. Specifically,*

$$\alpha = \kappa - \sum_{i=\kappa'+1}^{\kappa} \prod_{j=i}^{\kappa} (1 + \lfloor aj\rfloor + \lfloor b(1-j)\rfloor)\,.$$

Proof. By Lemma 1.2, $g(S((a,b))) = \lfloor a\alpha\rfloor$, for the greatest integer α where

$$\mathbb{N} \cap [b(\alpha-1), a\alpha]$$

is nonempty; in particular $\alpha \le \kappa$. The lower bound $\alpha \ge \kappa'$ is trivial when $\kappa' = 0$; if $b \le 1$ then

$$\kappa(a-1, b-1) = \left\lfloor \frac{b-1}{b-a} \right\rfloor \le 0,$$

and hence $\kappa' = 0$. Otherwise, $b > 1$ and so by Lemma 1.1,

$$\frac{\kappa' - 1}{\kappa'} \le \frac{a-1}{b-1};$$

rearranging we get $b(\kappa' - 1) \le a\kappa' - 1$. Hence the interval $[b(\kappa' - 1), a\kappa']$ has length at least 1. It must therefore contain an integer, so $\alpha \ge \kappa'$.

To prove the formula for α, for $i \le \kappa$ we define the function

$$f(i) = \begin{cases} 1 & \text{if } \alpha \le i, \\ 0 & \text{if } \alpha > i; \end{cases}$$

this gives us $\alpha = \kappa - \sum_{i=0}^{\kappa} f(i) = \kappa - \sum_{i=\kappa'+1}^{\kappa} f(i)$. We define f via $f(i) = \prod_{j=i}^{\kappa} \chi(j)$, for

$$\chi(j) = \begin{cases} 1 & \text{if } [b(j-1), aj] \cap \mathbb{N} \ne \varnothing, \\ 0 & \text{if } [b(j-1), aj] \cap \mathbb{N} = \varnothing. \end{cases}$$

We now have $\alpha = \kappa - \sum_{i=\kappa'+1}^{\kappa} \prod_{j=i}^{\kappa} \chi(j)$.

We now calculate $\chi(j)$ explicitly by showing that for $j \ge \kappa' + 1$, the interval $[b(j-1), aj]$ contains at most one integer. For $b \le 1$, we have $bj > aj \ge aj + (b-1)$ so $b(j-1) > aj - 1$. For $b > 1$, by Lemma 1.1 we have

$$\frac{a-1}{b-1} < \frac{\kappa'}{\kappa'+1} \le \frac{j-1}{(j-1)+1}, \qquad \text{for any } j - 1 \ge \kappa'.$$

Rearranging, we get $b(j-1) > aj - 1$. Hence $\big|[b(j-1), aj] \cap \mathbb{N}\big| \le 1$ and in fact $\chi(j)$ equals the number of integers in $[b(j-1), aj]$, that is,

$$\chi(j) = 1 + \lfloor aj \rfloor + \lfloor b(1-j) \rfloor. \qquad \square$$

We have $\alpha \in [\kappa', \kappa]$; in general, neither bound can be improved. The size of this interval, $\kappa - \kappa'$, can be arbitrarily large, when b/a is small. On the other hand, the following shows that $\kappa - \kappa'$ is small if $b/a > 2$. This is desirable, as it shortens the calculation for $g(S(I))$.

Proposition 3.2. *Let $0 < 2a < b$. Let κ, κ' be as in Theorem 3.1. Then*

$$\kappa - \kappa' = \begin{cases} 1 & \text{if } a < 1, \\ 0 & \text{if } a \ge 1. \end{cases}$$

Proof. For convenience, set

$$I = \Big(\frac{b-1}{b-a}, \frac{b}{b-a}\Big);$$

$\kappa - \kappa'$ counts the number of integers in I. Suppose first that $b \leq 1$. Then

$$\kappa(a-1, b-1) \leq 0,$$

so $\kappa' = 0$. Note that $b > 2a$ implies $b - a > a$, and hence $1/(b-a) < 1/a$, so $1 + a/(b-a) < 1 + a/a = 2$, and hence $\kappa = \lfloor 1 + a/(b-a) \rfloor = 1$. Suppose now that $a < 1 < b$. If $a \leq \frac{1}{2}$, then $b > 1 = \frac{1}{2} + a$. Alternatively, if $a > \frac{1}{2}$, then $b > 2a > \frac{1}{2} + a$. Hence $b > \frac{1}{2} + a$; rearranging we get $1/(b-a) < 2$. Hence I is of length less than 2, and can contain at most one integer. Therefore $\kappa - \kappa' \leq 1$. But I contains the integer $1 = (b-a)/(b-a)$, so $\kappa - \kappa' = 1$. Lastly, we consider the case $a \geq 1$. We have $b - 1 \geq b - a$, hence $(b-1)/(b-a) \geq 1$ and I does not contain 0 or 1. Suppose I contains integer $n \geq 2$. Then $2 \leq b/(b-a)$; rearranging we get $b \leq 2a$, a contradiction. Hence I contains no integers, and $\kappa - \kappa' = 0$. □

Computing $m(S((a, b)))$ is similar to computing $g(S((a, b)))$, in that we must count integers in an interval, only this time the intervals are open. We first prove a technical lemma, for which we recall Farey sequences (for an introduction see [Graham et al. 1994]). The nth Farey sequence F_n consists of all reduced fractions in $[0, 1]$ whose denominator is at most n, arranged in increasing order. The key property we require is that if a/b are c/d are consecutive terms in a Farey sequence, then $bc - ad = 1$.

Lemma 3.3. *Let $0 < a < b$. Let $n \in \mathbb{N}$ be minimal such that (an, bn) contains an integer. Suppose $n > 1$. Then (an, bn) contains exactly one integer.*

Proof. Suppose by way of contradiction that (an, bn) contains at least two integers. Then there is some $m \in \mathbb{N}$ such that $m, m+1 \in (an, bn)$. Set $d = \gcd(m, n)$. If $d > 1$ then $m/d \in (an/d, bn/d)$ violates the minimality of n. Similarly, $\gcd(m+1, n) = 1$. Let $m' \in (0, n-1)$ with $m = m' + kn$ for some integer k. We now consider the nth Farey sequence F_n. Both m'/n and $(m'+1)/n$ are elements of F_n; however $(m'+1)n - m'n = n > 1$, so they are not consecutive terms and there must be some p/q in F_n with $m'/n < p/q < (m'+1)/n$, with $q < n$. But then

$$\frac{m' + kn}{n} < \frac{p + qk}{q} < \frac{m' + 1 + kn}{n},$$

so $p + qk \in (aq, bq)$, violating the minimality of n. □

We now compute the multiplicity $m(S((a, b)))$. The reverse problem of finding an open interval whose semigroup possesses a given multiplicity, is solved in [Rosales and Vasco 2009]. A nondiscrete version is proved as Proposition 5 in [Rosales et al. 2003].

Theorem 3.4. *Suppose* $0 < a < b$. *Set* $\kappa'' = \kappa(1, b-a+1)$. *Then*

$$m(S((a,b))) = \lceil a\alpha \rceil,$$

where $\alpha \in \mathbb{N}$ *satisfies* $1 \le \alpha \le \kappa''$. *Specifically*,

$$\alpha = \sum_{i=0}^{\kappa''} \prod_{j=1}^{i} (2 + \lfloor aj \rfloor + \lfloor -bj \rfloor).$$

Proof. Set $m = m(S((a,b)))$, and let $\alpha \in \mathbb{N}$ be minimal such that $m/\alpha \in (a,b)$; then $m(S((a,b))) = \lceil a\alpha \rceil$. By Lemma 1.1, $1/(b-a+1) < \kappa''/(\kappa''+1)$. Rearranging, we find $\kappa''b - \kappa''a > 1$, so there is an integer $t \in (\kappa''a, \kappa''b)$. Suppose that $\alpha > \kappa''$. We then have $m/\alpha < m/\kappa'' \le t/\kappa''$; since m/α and t/κ'' are in (a,b), we conclude that $m/\kappa'' \in (a,b)$, which contradicts the minimality of α. Hence $\alpha \le \kappa''$.

We now prove the α formula. We proceed in a manner similar to Theorem 3.1, by defining

$$f(i) = \begin{cases} 1 & i \le \alpha, \\ 0 & i > \alpha, \end{cases}$$

via $f(i) = \prod_{j=1}^{i} (1 - \chi(j))$, where $\chi(j)$ is the number of integers in (aj, bj). For $i < \alpha$, $\chi(i) = 0$. By Lemma 3.3, $\chi(\alpha) = 1$, so $f(i) = 0$ for $i \ge \alpha$. Hence

$$\alpha = \sum_{i=0}^{\kappa''} f(i) = \sum_{i=0}^{\kappa''} \prod_{j=1}^{i} (1 - \chi(j)),$$

but $1 - \chi(j) = 2 + \lfloor aj \rfloor + \lfloor -bj \rfloor$. □

We have $\alpha \in [1, \kappa'']$; in general, neither bound can be improved. The upper bound κ'' can be arbitrarily large, when $b-a$ is small. On the other hand, the following shows that κ'' is small if $b-a$ is large, thus simplifying computation of m.

Proposition 3.5. *Let* $0 < a < b$. *Let* $n \in \mathbb{N}$ *be minimal with* $b - a > 1/n$. *Then* $\kappa'' = n$, *in the notation of Theorem 3.4.*

Proof. We have $1/n < b - a \le 1/(n-1)$, hence $n > 1/(b-a) \ge n-1$, so $\lfloor 1/(b-a) \rfloor = n-1$, and

$$\kappa'' = \left\lfloor \frac{b-a+1}{b-a} \right\rfloor = \left\lfloor 1 + \frac{1}{b-a} \right\rfloor = 1 + (n-1) = n.$$ □

Acknowledgments

The authors would like to thank the anonymous referee for extensive and helpful suggestions.

References

[Graham et al. 1994] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics*, 2nd ed., Addison-Wesley Publishing Company, Reading, MA, 1994. MR 97d:68003 Zbl 0836.00001

[Rosales and García-Sánchez 2009] J. C. Rosales and P. A. García-Sánchez, *Numerical semigroups*, Developments in Mathematics **20**, Springer, New York, 2009. MR 2010j:20091 Zbl 05623301

[Rosales and Urbano-Blanco 2006] J. C. Rosales and J. M. Urbano-Blanco, "Opened modular numerical semigroups", *J. Algebra* **306**:2 (2006), 368–377. MR 2007h:20063 Zbl 1109.20052

[Rosales and Vasco 2009] J. C. Rosales and P. Vasco, "Opened modular numerical semigroups with a given multiplicity", *Internat. J. Algebra Comput.* **19**:2 (2009), 235–246. MR 2010i:20082 Zbl 1187.20067

[Rosales et al. 2003] J. C. Rosales, P. A. García-Sánchez, J. I. García-García, and J. M. Urbano-Blanco, "Proportionally modular Diophantine inequalities", *J. Number Theory* **103**:2 (2003), 281–294. MR 2004k:20127 Zbl 1039.20036

[Rosales et al. 2005] J. C. Rosales, P. A. García-Sánchez, and J. M. Urbano-Blanco, "Modular Diophantine inequalities and numerical semigroups", *Pacific J. Math.* **218**:2 (2005), 379–398. MR 2007a:20056 Zbl 1184.20052

Received: 2010-06-30 Revised: 2010-09-29 Accepted: 2010-09-29

vadim@sciences.sdsu.edu *Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-7720, United States* http://www-rohan.sdsu.edu/~vadim/

bhappening@gmail.com *San Diego State University, San Diego, CA 92182-7720, United States*

INVOLVE 3:3(2010)

Distinct solution to a linear congruence

Donald Adams and Vadim Ponomarenko

(Communicated by Scott Chapman)

Given $n, k \in \mathbb{N}$ and $a_1, a_2, \ldots, a_k \in \mathbb{Z}_n$, we give conditions for the equation $a_1x_1 + a_2x_2 + \cdots + a_kx_k = 1$ in $\mathbb{Z}_n$ to admit solutions with all the x_i distinct.

A sufficient condition is that $k \leq \phi(n)$ and a_i be invertible in $\mathbb{Z}_n$ for all i.

If $n > 2$ is prime, the following conditions together are necessary and sufficient: $k \leq n$, each a_i is nonzero, and either $k < n$ or not all of the a_i are equal.

1. Linear congruence

Given $n, k \in \mathbb{N}$ and $a_1, a_2, \ldots, a_k \in \mathbb{Z}_n$, it is known classically [Uspensky and Heaslet 1939; Vandiver 1924] that the linear congruence

$$a_1x_1 + a_2x_2 + \cdots + a_kx_k = 1 \text{ (in } \mathbb{Z}_n) \tag{1}$$

has a solution if and only if $\gcd(a_1, a_2, \ldots, a_k) \in \mathbb{Z}_n^\times$, the group of units of $\mathbb{Z}_n$. We ask when such a solution exists with *distinct* $x_i \in \mathbb{Z}_n$, a question that appears to have been overlooked in the literature. In general, some additional conditions are necessary; for example, $1x_1 + 1x_2 + 1x_3 = 1$ does not have a solution with distinct $x_i \in \mathbb{Z}_3$.

Our partial solution has a stronger coefficient condition, and another restriction involving $\phi(n)$, the Euler totient. The general case remains open.

Theorem 1. *If $k \leq \phi(n)$ and $a_i \in \mathbb{Z}_n^\times$ $(1 \leq i \leq k)$, then there exist distinct $x_i \in \mathbb{Z}_n$ satisfying* (1).

Proof. We first construct $y_1, y_2, \ldots y_k$ iteratively, as will be explained. For notational convenience, for $i < j$ we set

$$y_{i,j} = y_i(1 - a_{i+1}y_{i+1})(1 - a_{i+2}y_{i+2}) \cdots (1 - a_{j-1}y_{j-1})$$

MSC2000: 11B50, 11D79.

Keywords: linear congruence, minimal zero-sum sequence, property B.

(note that $y_{i,i+1} = y_i$). We set $y_1 = a_1^{-1}$; for $j > 1$ we let y_j be any element chosen from $S_j \setminus T_j$, where

$$S_j = \{y \in \mathbb{Z}_n : 1 - a_j y \in \mathbb{Z}_n^\times\},$$
$$T_j = \{y \in \mathbb{Z}_n : y(1 + a_j y_{i,j}) = y_{i,j} \text{ for some } i \text{with } 1 \le i < j\}.$$

Note that the defining property of S_j ensures that $1 - a_j y_j$ is invertible, and that T_j ensures that $y_j \neq y_{i,j}(1 - a_j y_j) = y_{i,j+1}$, for all $i < j$.

Now, set $x_i = y_{i,k+1}$ for $1 \le i \le k$. Note that $a_1 x_1 + a_2 x_2 + \cdots + a_k x_k$ conveniently telescopes to 1, because $a_1 y_1 = 1$. Suppose that $x_i = x_j$ (for $i < j$). Then

$$y_{i,k+1} = y_{j,k+1}.$$

We may cancel the common terms, because they were constructed to be invertible, to get $y_{i,j+1} = y_{j,j+1} = y_j$, which contradicts our construction of y_j. Hence the x_i are distinct, and a solution to (1).

It remains to prove that $S_j \setminus T_j$ is nonempty. We first prove that

$$|S_j| = |\mathbb{Z}_n^\times| = \phi(n),$$

by showing that $f(y) = 1 - a_j y$ is a bijection on $\mathbb{Z}_n$, and thus $f(S_j) = \mathbb{Z}_n^\times$. If $f(y) = f(y')$, then $1 - a_j y = 1 - a_j y'$ and $a_j(y - y') = 0$, but a_j is invertible, hence $y = y'$. So f is injective on a finite set and hence bijective. Finally, we prove that $|T_j| \le j - 1 \le k - 1 < k \le \phi(n)$, by showing that $y(1 + a_j y_{i,j}) = y_{i,j}$ has at most one solution y. If $(1 + a_j y_{i,j})$ is invertible, then $y = (1 + a_j y_{i,j})^{-1} y_{i,j}$ is unique. If not, then there is some $m > 1$ with $m|n$ and $m|(1 + a_j y_{i,j})$. If there is a solution y then also $m|y_{i,j}$, so $m|(1 + a_j y_{i,j}) - a_j y_{i,j} = 1$, a contradiction. □

If n is prime, we can do better, solving the problem completely. Clearly it is necessary that $k \le n$, and that not all a_i are zero, that is, $\gcd(a_1, a_2, \ldots, a_k) \in \mathbb{Z}_n^\times$.

Theorem 2. *Let n be an odd prime, $k \le n$, and $\gcd(a_1, a_2, \ldots, a_k) \in \mathbb{Z}_n^\times$. Then there exist distinct $x_i \in \mathbb{Z}_n$ satisfying* (1), *if and only if either* (a) *$k < n$, or* (b) *not all of the a_i are equal.*

Proof. The nonzero a_i are in $\mathbb{Z}_n^\times$, and $\phi(n) = n - 1$, so unless there are n nonzero a_i, we can apply Theorem 1, and arbitrarily assign leftover distinct elements from $\mathbb{Z}_n$ to those x_i where $a_i = 0$. If $k = n$ and $a_1 = \cdots = a_k = t$, then there is only one possible solution, and it fails because $t(0 + 1 + \cdots + n) = tn(n + 1)/2 = 0$ in $\mathbb{Z}_n$.

Remaining is the case where $k = n$, the a_i are all nonzero and not all equal. Set $a_i' = a_i - a_1$. More than zero, but less than n, of the a_i' are nonzero, so we can find

distinct $x_i \in \mathbb{Z}_n$ with $a_1' x_1 + \cdots a_n' x_n = 1$. But now we have

$$\begin{aligned} a_1 x_1 + \cdots + a_n x_n &= (a_1' + a_1)x_1 + \cdots + (a_n' + a_1)x_n \\ &= (a_1' x_1 + \cdots + a_n' x_n) + a_1(x_1 + \cdots + x_n) \\ &= 1 + a_1(0 + 1 + \cdots + n) \\ &= 1 + a_1 n(n + 1/2) = 1 \quad \text{in } \mathbb{Z}_n. \end{aligned}$$

□

In fact, we believe that a similar result holds for composite n; this is supported by preliminary computer calculations. For example, consider $n = 6$, $k = 5$, $(a_1, a_2, a_3, a_4, a_5) = (2, 2, 2, 3, 3)$. Neither of the strong conditions of Theorem 1 are met; however $(x_1, x_2, x_3, x_4, x_5) = (2, 4, 5, 0, 1)$ satisfies (1).

Conjecture 3. Let $k < n$ and $\gcd(a_1, a_2, \ldots a_k) \in \mathbb{Z}_n^\times$. Then there exist distinct $x_i \in \mathbb{Z}_n$ satisfying (1).

2. Application

Fix the finite abelian group $\mathbb{Z}_n \times \mathbb{Z}_n$. We consider multisets[1] of elements such that their sum is zero; we call these zero-sum multisets. They have a rich literature and history [Geroldinger and Halter-Koch 2006], arising from fundamental number theoretic questions about nonunique factorization.

It is well known that the largest minimal (i.e. containing no other nontrivial zero-sum multiset) zero-sum multiset is of size $2n - 1$. Recently it has been shown [Gao et al. 2010] that any zero-sum multiset of this size contains some element of multiplicity $n - 1$. In [Gao and Geroldinger 2003] it was shown that the remaining multiplicities $a_1, a_2, \ldots a_k$ (where $a_1 + a_2 + \cdots + a_k = n$) must admit a solution to (1) in distinct elements of $\mathbb{Z}_n$, leaving open the question of when this occurs.

Corollary 4. *Let*

$$n > 0, \quad k \le \phi(n) \quad \text{and} \quad a_i \in \mathbb{N}, \quad \text{with} \begin{cases} a_1 + \cdots + a_k = n, \\ \gcd(a_i, n) = 1. \end{cases}$$

Then there is an irreducible zero-sum multiset in $\mathbb{Z}_n \times \mathbb{Z}_n$ whose elements have multiplicities $n - 1, a_1, a_2, \ldots, a_k$.

Corollary 5. *Let $n > 0$ be prime, $k \le n$, and $a_i \in \mathbb{N}$ with*

$$a_1 + \cdots + a_k = n \quad \text{and} \quad \gcd(a_1, a_2, \ldots, a_k, n) = 1.$$

Then there is an irreducible zero-sum multiset in $\mathbb{Z}_n \times \mathbb{Z}_n$ whose elements have multiplicities $n - 1, a_1, a_2, \ldots, a_k$ if and only if $1 < k < n$.

[1] For historical reasons these are called *sequences* in the literature, although the elements are not ordered.

References

[Gao and Geroldinger 2003] W. Gao and A. Geroldinger, "On zero-sum sequences in $\mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$", *Integers* **3** (2003), A8, 45 pp. MR 2004m:11015

[Gao et al. 2010] W. Gao, A. Geroldinger, and D. J. Grynkiewicz, "Inverse zero-sum problems. III", *Acta Arith.* **141**:2 (2010), 103–152. MR 2579841 Zbl 05691756

[Geroldinger and Halter-Koch 2006] A. Geroldinger and F. Halter-Koch, *Non-unique factorizations: Algebraic, combinatorial and analytic theory*, Pure and Applied Mathematics (Boca Raton) **278**, Chapman & Hall/CRC, Boca Raton, FL, 2006. MR 2006k:20001 Zbl 1113.11002

[Uspensky and Heaslet 1939] J. V. Uspensky and M. A. Heaslet, *Elementary Number Theory*, McGraw-Hill, New York, 1939.

[Vandiver 1924] H. S. Vandiver, "Discussions: on algorithms for the solution of the linear congruence", *Amer. Math. Monthly* **31**:3 (1924), 137–140. MR 1520388

Received: 2010-07-07 Revised: 2010-09-29 Accepted: 2010-09-29

DJUNIOR82@gmail.com *Arizona State University, Tempe, AZ 85287-1804, United States*

vadim@sciences.sdsu.edu *San Diego State University, Department of Mathematics and Statistics, 5500 Campanile Dr., San Diego, CA 92182-7720, United States*
http://www-rohan.sdsu.edu/~vadim/

INVOLVE 3:3(2010)

A note on nonresidually solvable hyperlinear one-relator groups

Jon P. Bannon and Nicholas Noblett

(Communicated by David R. Larson)

We prove that various nonresidually finite, nonresidually solvable groups of the form $\langle a, b \,|\, r^{r^w} = r^2 \rangle$ are sofic.

This paper concerns the sofic property discussed in the survey [Pestov 2008]. Particularly, we address Question 4.10 in that paper: the problem of Nate Brown asking whether or not every one-relator group is sofic. In [Bannon 2010], it is proved that the example in [Baumslag 1969] of a nonresidually finite nonresidually solvable one-relator group is a sofic group. The purpose of this paper is to exhibit more such examples in the following large class of nonresidually solvable one-relator groups introduced in [Baumslag et al. 2007]. Let $\mathbb{F}_2 = \langle a, b \,|\, \rangle$ denote the free group on two generators. Let $r, w \in \mathbb{F}_2$ be two elements that do not commute. In [Baumslag et al. 2007], the authors show that the group

$$\Gamma_{r,w} = \langle a, b \,|\, r^{r^w} = r^2 \rangle = \langle a, b \,|\, r = [r, (r^{-1})^w] \rangle$$

has the same finite quotients as the group

$$\langle a, b \,|\, r \rangle,$$

and is therefore not residually finite. We point out that none of the groups $\Gamma_{r,w}$ are residually solvable, since $r = [r, (r^{-1})^w]$ lies in every derived subgroup of $\Gamma_{r,w}$. In [Bannon 2010], it is shown that the group $\Gamma_{ab,a}$ is sofic. The proof in [Bannon 2010] uses [Dykema 2010, Corollary 3.4], that HNN extensions of sofic groups over amenable subgroups remain sofic. The proof in [Bannon 2010] uses the fact that $\Gamma_{ab,a}$ is an HNN extension of an amenable one-relator group. We shall extend this result to certain other of the groups $\Gamma_{r,w}$. If r and w generate $\mathbb{F}_2$, then $\Gamma_{r,w}$ embeds naturally as a subgroup of $\Gamma_{ab,a}$, and since the sofic property passes to subgroups, $\Gamma_{r,w}$ is sofic. The first result of this short note is that there exist r, w that do not generate $\mathbb{F}_2$, yet the group $\Gamma_{r,w}$ is sofic. More precisely, we prove:

MSC2000: primary 46L10; secondary 20F65.
Keywords: mathematics.
Noblett is an undergraduate at Siena College, Loudonville, NY.

Theorem 1. *The group $\Gamma_{a,b^{-1}ab}$ is sofic.*

Proof. Since $\Gamma_{a,b^{-1}ab} = \langle a, b \,|\, (bab^{-1})^{-2}a^{-1}(bab^{-1})^{-1}a(bab^{-1})a^{-1}(bab^{-1})a\rangle$, following [McCool and Schupp 1973], we let $a_0 = a$ and $a_{-1} = bab^{-1}$ and realize $\Gamma_{a,b^{-1}ab}$ as the HNN extension

$$\langle a_0, a_{-1}, t \,|\, (a_{-1})^{-2}a_0^{-1}(a_{-1})^{-1}a_0a_{-1}a_0^{-1}a_{-1}a_0, t^{-1}a_{-1}t = a_0\rangle$$

of the one-relator group $H_1 = \langle a_0, a_{-1} \,|\, a_0(a_{-1})^{-2}a_0^{-1}(a_{-1})^{-1}a_0a_{-1}a_0^{-1}a_{-1}\rangle$, where by the Freiheitssatz $\langle a_{-1}\rangle$ and $\langle a_0\rangle$ are copies of $\mathbb{Z}$ which in the HNN extension we identify by identifying a_{-1} with a_0. Letting $b_1 = a_0a_{-1}a_0^{-1}$ and $b_0 = a_{-1}$ we may identify H_1 as the HNN extension

$$\langle b_0, b_1, s \,|\, b_1^{-2}b_0^{-1}b_1b_0,\ s^{-1}b_1s = b_0\rangle$$

of the one-relator group $H_2 = \langle b_0, b_1, s \,|\, b_1^{-2}b_0^{-1}b_1b_0\rangle$, where we identify the two copies $\langle b_0\rangle$ and $\langle b_1\rangle$ of $\mathbb{Z}$ as above. By [Ceccherini-Silberstein and Grigorchuk 1997], the group H_2 is amenable, and hence by the argument in [Bannon 2010], the group H_1 is sofic. Since $\Gamma_{a,b^{-1}ab}$ is an HNN extension of a sofic group with respect to identified copies of the amenable group $\mathbb{Z}$, it follows that $\Gamma_{a,b^{-1}ab}$ is sofic. □

In this proof we used in an essential way that the identified subgroups are amenable and therefore invoke the full hypotheses of Corollary 3.4 of [Dykema 2010], whereas in [Bannon 2010], the group $\Gamma_{ab,a}$ is an HNN extension of an amenable group and so any pair of identified subgroups would work. We next illustrate that there are groups of the form $\Gamma_{r,w}$ that do not in an obvious way fall to the method of [Bannon 2010].

Theorem 2. *The group $\Gamma_{a,b^2} = \langle a, b \,|\, a = [a, (a^{-1})^{b^2}]\rangle$ is isomorphic to*

$$(G * \mathbb{Z}) *_{\mathbb{F}_2},$$

where G is a one-relator amenable group.

Proof. Since $\Gamma_{a,b^2} = \langle a, b \,|\, a^{-2}(b^2ab^{-2})a(b^2ab^{-2})^{-1}\rangle$, then letting $a_0 = a$ and $a_{-2} = b^2ab^{-2}$ we have that Γ_{a,b^2} is isomorphic to the HNN extension

$$\langle a_0, a_{-1}, a_{-2}, t \,|\, a_0^{-2}a_{-2}a_0(a_{-2})^{-1}, t^{-1}a_{-2}t = a_{-1}, t^{-1}a_{-1}t = a_0\rangle$$

of the one-relator group $\langle a_0, a_{-1}, a_{-2} \,|\, a_0^{-2}a_{-2}a_0(a_{-2})^{-1}\rangle$, with the isomorphism from the free subgroup $\langle a_{-2}, a_{-1}\rangle$ with $\langle a_{-1}, a_0\rangle$ extending the set map that sends a_{-2} to a_{-1} and a_{-1} to a_0. But the relator $a_0^{-2}a_{-2}a_0(a_{-2})^{-1}$ does not involve a_{-1}, hence $\langle a_0, a_{-1}, a_{-2} \,|\, a_0^{-2}a_{-2}a_0(a_{-2})^{-1}\rangle = \langle a_{-1}\rangle * \langle a_0, a_{-2} \,|\, a_0^{-2}a_{-2}a_0(a_{-2})^{-1}\rangle$. □

References

[Bannon 2010] J. Bannon, "A non-residually solvable hyperlinear one-relator group", preprint, 2010, Available at http://tinyurl.com/BannonOneRelator.

[Baumslag 1969] G. Baumslag, "A non-cyclic one-relator group all of whose finite quotients are cyclic", *J. Austral. Math. Soc.* **10** (1969), 497–498. MR 40 #7337 Zbl 0214.27402

[Baumslag et al. 2007] G. Baumslag, C. F. Miller, III, and D. Troeger, "Reflections on the residual finiteness of one-relator groups", *Groups Geom. Dyn.* **1**:3 (2007), 209–219. MR 2008d:20056 Zbl 1141.20024

[Ceccherini-Silberstein and Grigorchuk 1997] T. G. Ceccherini-Silberstein and R. I. Grigorchuk, "Amenability and growth of one-relator groups", *Enseign. Math.* (2) **43**:3-4 (1997), 337–354. MR 99b:20057 Zbl 0897.20022

[Dykema 2010] K. J. Dykema, "Free products of sofic groups with amalgamation over amenable groups", preprint, 2010. arXiv 1003.1675v

[McCool and Schupp 1973] J. McCool and P. E. Schupp, "On one relator groups and HNN extensions", *J. Austral. Math. Soc.* **16** (1973), 249–256. MR 49 #2952 Zbl 0288.20046

[Pestov 2008] V. G. Pestov, "Hyperlinear and sofic groups: a brief guide", *Bull. Symbolic Logic* **14**:4 (2008), 449–480. MR 2009k:20103 Zbl 05495887

Received: 2010-07-26 Revised: Accepted: 2010-07-26

jbannon@siena.edu *Siena College, Department of Mathematics, Loudonville, NY 12211, United States*

nb11nobl@siena.edu *Siena College, Department of Mathematics, Loudonville, NY 12211, 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.

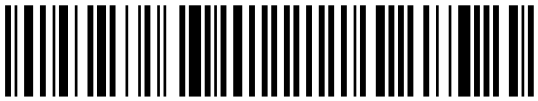
Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

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. 3



1944-4176(2010)3:3;1-E