involve

a journal of mathematics

msp

2016 vol. 9, no. 5

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2016 is US $160/year for the electronic version, and $215/year (+$35, if 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 MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o 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

nonprofit scientific publishing

http://msp.org/

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.721

An iterative strategy for Lights Out on Petersen graphs

Bruce Torrence and Robert Torrence

(Communicated by Kenneth S. Berenhaut)

We establish some preliminary results for Sutner's σ^+ game, known as Lights Out, played on the generalized Petersen graph $P(n, k)$. While all regular Petersen graphs admit game configurations that are not solvable, we prove that every game on the $P(2n, n)$ graph has a unique solution. Moreover, we introduce a simple iterative strategy for finding the solution to any game on $P(2n, n)$, and generalize its application to a wider class of graphs.

Background

All graphs are assumed to be undirected, without loops or multiple edges. The σ^+ game is a well-known single-player game that can be played on any graph [Sutner 1990]. The handheld game *Lights Out*, released by Tiger Electronics in 1993, was the σ^+ game on a 5×5 grid graph. Since then, the name Lights Out has been widely used and is synonymous with the σ^+ game.

The idea of the game is simple. Each vertex is in one of two states: on or off (think of the vertices as lights). Each vertex also acts as a button. When the player pushes a button, its state toggles, and so do the states of each of its neighbors. Given a graph and an initial configuration of lit vertices, the goal is to turn out all the lights. Several playable versions of the game can be found online [Antonick 2013; Scherphuis 2012; Torrence 2016].

Lights Out has been extensively studied on grid graphs, and a wide range of generalizations have been explored. Our goal here is to present some preliminary results for Lights Out played on the generalized Petersen graphs. In the process, we introduce an iterative strategy that can successfully solve games on a larger class of graphs.

MSC2010: 05C50, 37B99, 90C27, 91A43, 05C57.
Keywords: Lights Out, Petersen graph, game, sigma plus game.

Definitions

Given a graph G with finite vertex set $V(G)$ and edge set $E(G)$, label the vertices $v_1, v_2, \ldots, v_n$. The *adjacency matrix* $\mathrm{Adj}(G)$ is the $n \times n$ symmetric matrix with a one in position (i, j) if $v_i v_j \in E(G)$, and a zero otherwise. We let I_n denote the $n \times n$ identity matrix.

A *lights out configuration* or *game* $\boldsymbol{g}$ is an $n \times 1$ column vector with entries in the two-element field $\mathbb{F}_2$. If there is a one in position i, we say that v_i is *lit*; if there is a zero we say it is *off*. We let $\mathbf{0}$ and $\mathbf{1}$ denote the all-off and all-on configurations, respectively.

Similarly, a *lights out strategy* $\boldsymbol{s}$ is an $n \times 1$ column vector with entries in $\mathbb{F}_2$. It denotes a collection of buttons to be pushed; if there is a one in position i, we say that vertex v_i is *pushed*. The vector $\mathbf{1}$ represents the strategy where every button is pushed.

If one begins with the all-off configuration and invokes strategy $\boldsymbol{s}$, the resulting configuration is the matrix product $(\mathrm{Adj}(G) + I_n)\boldsymbol{s}$, with arithmetic carried out modulo 2 [Sutner 1990]. Since the matrix $\mathrm{Adj}(G) + I_n$ is so important, we call it the *transition matrix* for the σ^+ game on G, and denote it $A(G)$, or simply A if the underlying graph is understood.

In general, if one begins with configuration $\boldsymbol{g}$ and invokes strategy $\boldsymbol{s}$, the configuration that results is $A\boldsymbol{s} + \boldsymbol{g}$, with arithmetic carried out modulo 2. We say that strategy $\boldsymbol{s}$ *solves* game $\boldsymbol{g}$ if $A\boldsymbol{s} + \boldsymbol{g} = \mathbf{0}$, the all-off configuration. Since we are working modulo 2, this is equivalent to saying $A\boldsymbol{s} = \boldsymbol{g}$. We say configuration $\boldsymbol{g}$ is *solvable* if there exists a strategy that solves it.

For any vector $\boldsymbol{g} \in (\mathbb{F}_2)^n$, the *light number* of $\boldsymbol{g}$ is the sum of the entries in $\boldsymbol{g}$. The *parity of* $\boldsymbol{g}$ is odd or even according to the parity of the light number of $\boldsymbol{g}$.

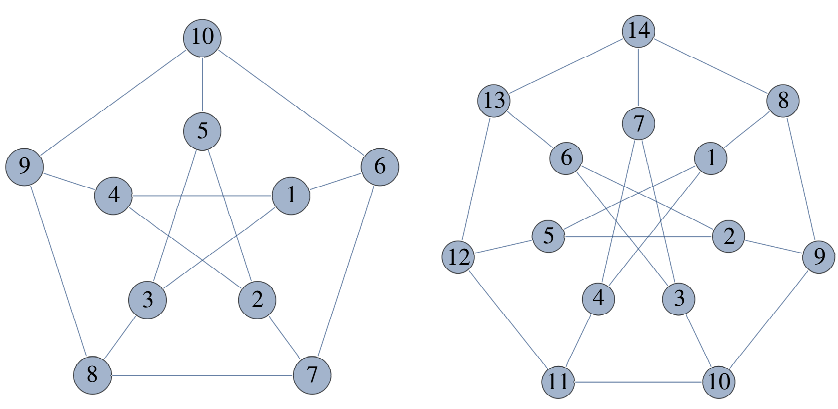


Figure 1. The Petersen graphs $P(5, 2)$ and $P(7, 3)$.

Let n and k be integers with $n \geq 3$ and $1 \leq k < n$. The *generalized Petersen graph* $P(n, k)$ is a graph with $2n$ vertices arranged in two concentric "rings" with n vertices in each ring. We label the vertices on the inner (or lower) ring v_1 to v_n, and on the outer (or upper) ring v_{n+1} to v_{2n}. Each outer-ring vertex v_i is connected to the inner-ring vertex v_{i-n}, and to its two nearest outer-ring neighbors (v_{i-1} and v_{i+1} for $n+1 < i < 2n$, while v_{2n} is connected to v_{2n-1} and v_{n+1}). Each inner-ring vertex v_i is connected to v_{i+k}, where the index $i+k$ is reduced modulo n if $i+k > n$. In other words, vertices on the inner ring are connected to one another by "skipping" k vertices. The classic Petersen graph $P(5, 2)$ is shown on the left in Figure 1, with $P(7, 3)$ beside it.

For a given value of n, one need only consider $k = 1$ through $k = \lfloor n/2 \rfloor$, since the graphs $P(n, k)$ and $P(n, n-k)$ are isomorphic. (The first skips k vertices clockwise, the second skips k vertices counterclockwise.) Also, $P(2n, n)$ is not regular, as the inner vertices have valence 2 (see Figure 2). All other Petersen graphs are 3-regular.

An oscillating strategy

The Petersen graphs have an important property: if one refers to the vertices $\{v_1, \ldots, v_n\}$ as the lower vertices, and $\{v_{n+1}, \ldots, v_{2n}\}$ as the upper vertices, then for each i with $1 \leq i \leq n$, there is an edge $v_i v_{n+i}$, and these are the only edges between a lower vertex and an upper vertex. In other words, the adjacency matrix $\mathrm{Adj}(P(n, k))$ has the block form

$$\begin{pmatrix} C' & I_n \\ I_n & D' \end{pmatrix},$$

where C' is the adjacency matrix for the subgraph induced by the lower vertices, D' is the adjacency matrix for the subgraph induced by the upper vertices, and where the two identity matrices specify the edges $v_i v_{n+i}$ between upper and lower vertices.

We now introduce an iterative "oscillating" lights out strategy that can be applied to any graph whose adjacency matrix has this structure. That is, in this section we suppose that we are given a graph G with $2n$ vertices such that when the vertices of G are appropriately ordered its adjacency matrix has the block form above, with I_n in the lower-left and upper-right corners (and where C' and D' can be any symmetric matrices over $\mathbb{F}_2$ with zeros on their respective main diagonals). We call such a graph *Petersen-like*.

The strategy works as follows: Suppose that $\boldsymbol{g}$ is a game on the Petersen-like graph G. For each lower vertex v_k that is lit, push button v_{n+k}. This will have the effect of turning off all lower vertices. Then, for each upper vertex v_{n+k} that is now lit, push button v_k. This will have the effect of turning off all upper vertices (and possibly lighting or relighting some lower ones). Together, we call these two operations performed in succession "one oscillation". One then repeats the process: After the first oscillation, if any lower vertices are lit, push the corresponding upper vertices. Then if any upper vertices are still lit, push the corresponding lower vertices, and so on.

The strategy can be expressed explicitly in matrix form. Let U and L be the $2n \times 2n$ matrices, defined in block form as

$$U = \begin{pmatrix} 0 & 0 \\ I_n & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & I_n \\ 0 & 0 \end{pmatrix},$$

where 0 denotes the $n \times n$ zero matrix. Then for any configuration $\boldsymbol{g}$, the upper vertices that correspond to lit lower vertices are $U\boldsymbol{g}$, so the configuration that results after pushing the upper vertices corresponding to the lit lower vertices is

$$\boldsymbol{g}_1 = AU\boldsymbol{g} + \boldsymbol{g} = (AU + I)\boldsymbol{g}.$$

And pushing the lower vertices corresponding to lit upper vertices in $\boldsymbol{g}_1$ yields the configuration

$$\boldsymbol{g}_2 = AL\boldsymbol{g}_1 + \boldsymbol{g}_1 = (AL + I)\boldsymbol{g}_1 = (AL + I)(AU + I)\boldsymbol{g}.$$

It follows that after performing k successive upper-lower iterations, the final configuration is

$$[(AL + I)(AU + I)]^k \boldsymbol{g}.$$

We call the matrix $[(AL + I)(AU + I)]$ the *oscillating matrix.*

Noting that the lights out transition matrix A has the form

$$A = \begin{pmatrix} C & I \\ I & D \end{pmatrix},$$

where $C = C' + I$ and $D = D' + I$, it is simple to calculate the oscillating matrix:

$$(AL + I) = \begin{pmatrix} C & I \\ I & D \end{pmatrix} \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & C \\ 0 & I \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & C \\ 0 & 0 \end{pmatrix}.$$

Similarly,

$$(AU + I) = \begin{pmatrix} C & I \\ I & D \end{pmatrix} \begin{pmatrix} 0 & 0 \\ I & 0 \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 0 \\ D & 0 \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ D & I \end{pmatrix}.$$

The product is

$$(AL + I)(AU + I) = \begin{pmatrix} I & C \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ D & I \end{pmatrix} = \begin{pmatrix} CD & C \\ 0 & 0 \end{pmatrix}.$$

Theorem 1. *Suppose G is a graph with $2n$ vertices whose lights out transition matrix has the form*

$$A = \begin{pmatrix} C & I_n \\ I_n & D \end{pmatrix}.$$

If the product CD is nilpotent, then A is invertible, and the oscillating strategy will solve any initial configuration on G. Specifically, if $(CD)^m$ is the zero matrix, then at most $m+1$ oscillations are required.

Proof. A straightforward inductive argument shows that for $k \geq 1$,

$$[(AL+I)(AU+I)]^k = \begin{pmatrix} (CD)^k & (CD)^{k-1}C \\ 0 & 0 \end{pmatrix}.$$

So if $(CD)^m = 0$ and if $\boldsymbol{g}$ is any initial configuration, $[(AL+I)(AU+I)]^{m+1}\boldsymbol{g} = \boldsymbol{0}$, and we see that $m+1$ iterations will suffice to solve $\boldsymbol{g}$.

It remains to show that A is invertible. Since we have shown that the oscillating strategy will solve any configuration $\boldsymbol{g}$, and since there are 2^{2n} possible configurations and the same number of possible strategies, we see that for each game there is precisely one strategy that solves it. So $A\boldsymbol{s} = \boldsymbol{g}$ has a unique solution $\boldsymbol{s}$ for each game $\boldsymbol{g}$. This means that A is invertible. $\square$

Examples

In each example, G is a Petersen-like graph with $2n$ vertices and with lights out transition matrix $A = \left(\begin{smallmatrix} C & I_n \\ I_n & D \end{smallmatrix}\right)$.

Example 2. Suppose n is even, $C = I_n$, and D is the all-ones matrix. Since n is even and we are working over $\mathbb{F}_2$, one has $(CD)^2 = D^2 = 0$ (the zero matrix). In graph theoretic terms, the lower vertices have no edges among them, while the upper vertices induce the complete graph K_n. According to Theorem 1, at most three iterations of the oscillating strategy will solve any game.

Note that playing Lights Out on K_n itself is quite boring. Pushing any single vertex changes the state of *every* vertex. Hence the only solvable games are $\boldsymbol{0}$ and $\boldsymbol{1}$. It is interesting that adding a "dead end" edge to each vertex of K_n transforms it into a graph where every game is solvable, and where the oscillating strategy can solve any game.

Example 3. Suppose m is a natural number and $n = 2m$. Let

$$C = \begin{pmatrix} I_m & I_m \\ I_m & I_m \end{pmatrix}, \quad D = \begin{pmatrix} T & S \\ S & T \end{pmatrix},$$

where S and T may be any $m \times m$ symmetric matrices over $\mathbb{F}_2$ such that T has ones along its main diagonal. Then in block form, each of the four blocks of CD

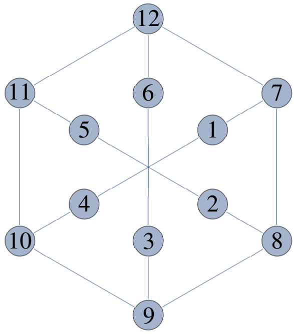


Figure 2. The Petersen graph $P(6, 3)$.

is $S + T$, and since we are working over $\mathbb{F}_2$, we know $(CD)^2$ is the zero matrix. According to Theorem 1, at most three iterations suffice to solve any game.

If $T = I_m$ and $S = 0$, then G is isomorphic to m copies of the path graph P_4. In this setting, the upper vertices have valence 1, and the lower vertices have valence 2.

If T is the $m \times m$ tridiagonal matrix, and S is the $m \times m$ matrix with ones in the upper-right and lower-left corners only, then G is the Petersen graph $P(2m, m)$. See Figure 2.

Corollary 4. *For every natural number m, the transition matrix for the Petersen graph $P(2m, m)$ is invertible, and at most three iterations of the oscillating strategy suffice to solve any game.* □

Example 5. Suppose n is even. Let $C = (c_{i,j})$ be the $n \times n$ matrix where $c_{1,1} = 1$, all other entries in the first row and column are zero, and if $i > 1$ and $j > 1$, then $c_{i,j} = 1$. In graph theoretic terms, the lower vertices induce the graph that is the union of the isolated vertex v_1 with the complete graph on $v_2, \ldots, v_n$. Let D be the $n \times n$ matrix with ones in the first row and column, ones on the main diagonal, and zeros elsewhere. In graph theoretic terms, the upper vertices induce the star graph $K_{1,n-1}$ with hub at v_{n+1}. It is a simple matter to check that $C = C^2$, so C is not nilpotent. Also, $D^n = D^2$ is the matrix obtained from D by swapping all ones with zeros and zeros with ones, so D is not nilpotent. Yet one readily verifies that $(CD)^2$ is the zero matrix, so Theorem 1 applies: every game on G is solvable, and at most three iterations of the oscillating strategy suffice to solve any game.

Example 6. Suppose n is even. Let $C = (c_{i,j})$ be the $n \times n$ matrix where $c_{1,1} = c_{2,2} = 1$, all other entries in the first two rows and columns are zero, and for either $i > 2$ or $j > 2$, we have $c_{i,j} = 1$. In graph theoretic terms, the lower vertices induce the graph that is the union of isolated vertices v_1 and v_2 with the complete graph on $v_3, \ldots, v_n$. Let D be the $n \times n$ matrix with ones in the first row and column, ones on the main diagonal, and zeros elsewhere (exactly the same as it was in

Example 5). In graph theoretic terms, the upper vertices induce the star graph $K_{1,n-1}$ with hub at v_{n+1}. It is a simple matter to check that $C^n = C^2$ has precisely two ones, in positions $(1, 1)$ and $(2, 2)$, so C is not nilpotent. Also, $D^n = D^2$ is the matrix obtained from D by swapping all ones with zeros and zeros with ones, so D is not nilpotent. Yet one readily verifies that $(CD)^3$ is the zero matrix, so Theorem 1 applies: every game on G is solvable, and at most four iterations of the oscillating strategy suffice to solve any game.

These examples should make clear that Theorem 1 is widely applicable. In particular, the last two examples show that for every even n there are multiple Petersen-like graphs for which neither C nor D is nilpotent, but CD is.

A restriction

In each example from the previous section, G is a Petersen-like graph where Theorem 1 applies: G has σ^+ transition matrix $A = \left(\begin{smallmatrix} C & I_n \\ I_n & D \end{smallmatrix}\right)$, and the product CD is nilpotent. Observe that in each example, C and D are $n \times n$ matrices where n is even. This is no accident.

Theorem 7. *Suppose C and D are symmetric $n \times n$ matrices over a field of characteristic* 2, *with ones along their main diagonals. If CD is nilpotent*, *then n is even.*

Before proving this, we present a lemma:

Lemma 8. *Suppose $C = (c_{i,j})$ and $D = (d_{i,j})$ are symmetric $n \times n$ matrices over a field of characteristic* 2, *with ones along their main diagonals. Then the trace* $\operatorname{tr}(CD) = 0$ *if and only if n is even.*

Proof of Lemma 8. The k-th diagonal entry of CD is

$$c_{k,1}d_{1,k} + c_{k,2}d_{2,k} + \cdots + c_{k,n}d_{n,k} = c_{1,k}d_{1,k} + c_{2,k}d_{2,k} + \cdots + c_{n,k}d_{n,k}$$

since C is symmetric. The expression on the right is the sum of the entries in the k-th column of the Hadamard product $C \circ D$. Summing over all columns k, we see that $\operatorname{tr}(CD)$ is the sum of *all* entries of $C \circ D$. But $C \circ D$ is symmetric, since C and D are, and has ones on its main diagonal since each of C and D do. Therefore, the nondiagonal entries appear in pairs (and so cancel modulo 2), and the sum of the diagonal entries is n. So over a field of characteristic 2, we have $\operatorname{tr}(CD) \equiv n \pmod 2$. □

Proof of Theorem 7. We prove the contrapositive. Suppose n is odd. The characteristic polynomial of an $n \times n$ matrix M over a field of characteristic 2 has $\operatorname{tr}(M) \pmod 2$ as the coefficient to the term with power $n - 1$. Since n is odd, the lemma says that $\operatorname{tr}(CD)$ is nonzero. Therefore the characteristic polynomial of CD has a nonzero coefficient for the term with power $n - 1$. This means that CD has a nonzero eigenvalue, and therefore cannot be nilpotent. □

Results for Petersen graphs

The transition matrix $A(G)$ governs the behavior of the σ^+ game on G, and its nullity is particularly important. If G has n vertices, there are 2^n configurations on G. If $A(G)$ has nullity k, then there are 2^k nullspace vectors in $(\mathbb{F}_2)^n$. This means there are 2^{n-k} solvable games on G, and every solvable game has 2^k distinct strategies that solve it: if strategy $\boldsymbol{s}$ solves game $\boldsymbol{g}$, so does $\boldsymbol{s}+\boldsymbol{n} \in (\mathbb{F}_2)^n$ for each nullspace vector $\boldsymbol{n}$. If $A(G)$ is nonsingular, then every game is solvable and has a unique solution.

Our first goal is to determine which Petersen graphs have nonsingular transition matrices.

Lemma 9. *If G is a graph where every vertex is odd-valent, then the σ^+ transition matrix $A(G)$ is singular. In particular, the all-on strategy $\mathbf{1}$ is in the nullspace of $A(G)$.*

Proof. Suppose every vertex in G is odd-valent. Consider the strategy $\mathbf{1}$, where every button is pushed once. Let v be a button. Since v gets pushed, it changes state once on that account. But v has odd valence, so in addition it will change state an odd number of times (once for each button adjacent to it). So ultimately v changes state an even number of times, and hence is left unchanged. Therefore $\mathbf{1}$ is in the nullspace of $A(G)$, so $A(G)$ must be singular. □

Corollary 10. *If G is a graph where every vertex is odd-valent, and $\boldsymbol{s}$ is a strategy that solves the σ^+ game $\boldsymbol{g}$, then the complementary strategy $\mathbf{1}-\boldsymbol{s}$ also solves game $\boldsymbol{g}$.*

Proof. Suppose $\boldsymbol{s}$ solves game $\boldsymbol{g}$. This means $A\boldsymbol{s}=\boldsymbol{g}$, where $A=A(G)$ is the transition matrix for G. Since we are working modulo 2, we know $\mathbf{1}-\boldsymbol{s}=\mathbf{1}+\boldsymbol{s}$, and we have

$$A(\mathbf{1}-\boldsymbol{s}) = A\mathbf{1}+A\boldsymbol{s} = A\boldsymbol{s} = \boldsymbol{g},$$

since $A\mathbf{1}=\mathbf{0}$ by Lemma 9. So strategy $\mathbf{1}-\boldsymbol{s}$ also solves game $\boldsymbol{g}$. □

We now focus on Petersen graphs. Corollary 4 tells us that the transition matrix for $P(2n,n)$ is invertible, and that the oscillating strategy can be used to solve any game on this graph. Note that for graphs of this type, the lower vertices have valence 2, while the upper vertices have valence 3.

Theorem 11. *The generalized Petersen graph $P(n,k)$ has nonsingular σ^+ transition matrix if and only if $n=2k$.*

Proof. If $n \neq 2k$, then $P(n,k)$ is 3-regular, so by Lemma 9 its transition matrix is singular. If $n=2k$, then Corollary 4 guarantees that the transition matrix is nonsingular. □

While we have shown that the nullity of A is strictly positive for the regular Petersen graphs, determining its precise value is a subtle business (see Table 1). Both Sutner [1988], and Anderson and Feil [1998] pointed out a similar situation for grid graphs, and much work was done subsequently to make sense of it [Barua and Ramakrishnan 1996; Goldwasser et al. 1997; Sutner 2000].

Lemma 12. *Let $\boldsymbol{g}$ and $\boldsymbol{h}$ be vectors in $(\mathbb{F}_2)^n$. Then $\boldsymbol{g}+\boldsymbol{h} \in (\mathbb{F}_2)^n$ has even parity if and only if $\boldsymbol{g}$ and $\boldsymbol{h}$ have the same parity.*

Proof. Let m and n be the light numbers of $\boldsymbol{g}$ and $\boldsymbol{h}$. Let k be the number of coordinate positions where $\boldsymbol{g}$ and $\boldsymbol{h}$ both have the value 1. Then the light number of $\boldsymbol{g}+\boldsymbol{h}$ is

$$(m-k)+(n-k)=m+n-2k \equiv m+n \pmod 2.$$

But $m+n$ is even if and only if m and n have the same parity. □

Lemma 13. *If G is a graph where every vertex is odd-valent, and $\boldsymbol{g}$ is a solvable game, then $\boldsymbol{g}$ is even.*

Proof. Let $A=A(G)$ be the transition matrix for G. Observe that A has an even number of ones in each column, since $A=\mathrm{Adj}(G)+I$, and the adjacency matrix for G has an odd number of ones in each column and zeros on the diagonal. Now for any strategy $\boldsymbol{s}$, the vector $A\boldsymbol{s}$ is the simply the sum of those columns of A where $\boldsymbol{s}$ has a one. It follows from Lemma 12 that any such sum has even parity. So if $A\boldsymbol{s}=\boldsymbol{g}$, it must be the case that $\boldsymbol{g}$ has even parity. □

Corollary 14. *If G is a graph where every vertex is odd-valent, and the nullity of $A(G)$ is 1, then the solvable games are precisely the games with an even number of lights lit. Every solvable game has precisely two solutions, and they are complements of each other.*

Proof. On any graph, precisely half of all possible games have an even light number. If the nullity of the transition matrix for any graph is 1, precisely half of all games are solvable (and every solvable game has two solutions). But Lemma 13 says that on an odd-valent graph, every solvable game has an even light number, hence these two sets (solvable games and games with an even light number) must agree. If strategy $\boldsymbol{s}$ solves game $\boldsymbol{g}$, Corollary 10 guarantees the complement $\mathbf{1}-\boldsymbol{s}$ is the other solution. □

Table 1 shows that among the various Petersen graphs, there are many examples satisfying the assumptions of Corollary 14 (e.g., $P(n,1)$ when n is odd). Other questions and conjectures are suggested by the table. For example, it seems natural to conjecture that the parity of n matches the parity of the nullity of $A(P(n,k))$ for all k.

$n\downarrow$ $k\rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1																			
4	4	0																		
5	1	5																		
6	2	2	0																	
7	1	1	1																	
8	4	2	4	0																
9	1	1	1	1																
10	2	6	6	2	0															
11	1	1	1	1	1															
12	4	2	4	2	4	0														
13	1	1	1	1	1	1														
14	2	2	2	2	2	2	0													
15	1	5	5	9	1	1	5													
16	4	2	4	2	4	2	4	0												
17	1	1	1	1	9	1	9	1												
18	2	2	2	2	2	2	2	2	0											
19	1	1	1	1	1	1	1	1	1											
20	4	6	8	2	4	2	8	6	4	0										
21	1	1	1	1	1	1	1	1	1	1										
22	2	2	2	2	2	2	2	2	2	2	0									
23	1	1	1	1	1	1	1	1	1	1	1									
24	4	2	4	2	4	2	4	2	4	2	4	0								
25	1	5	5	1	1	1	5	5	1	1	1	5								
26	2	2	2	2	2	2	2	2	2	2	2	2	0							
27	1	1	1	1	1	1	1	1	1	1	1	1	1							
28	4	2	4	2	4	2	4	2	4	2	4	2	4	0						
29	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
30	2	6	6	10	2	2	6	6	2	2	10	6	6	2	0					
31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1					
32	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	0				
33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
34	2	2	2	2	10	2	10	2	2	10	2	10	2	2	2	2	0			
35	1	5	5	1	1	1	5	5	1	1	1	5	5	1	1	1	5			
36	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	0		
37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
38	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
40	4	6	8	2	4	2	8	6	4	2	4	6	8	2	4	2	8	6	4	0

Table 1. Nullity of $A(P(n,k))$ for $n \leq 40$.

Graph	Oscillating strategy
$P(4, 1)$	L
$P(5, 2)$	L
$P(6, 1)$	LUL
$P(6, 2)$	UL
$P(9, 3)$	LUL
$P(12, 1)$	$LULUL$
$P(12, 3)$	$LULULULU$
$P(12, 5)$	$LULUL$
$P(18, 3)$	$LULUL$
$P(18, 6)$	$LULU$
$P(36, 3)$	$LULULULUL$
$P(36, 15)$	$LULULULUL$

Table 2. Regular Petersen graphs $P(n, k)$ with $n \leq 72$ where an oscillating strategy suffices to solve any solvable game. Read each strategy from left to right. For example, UL means first push the upper buttons (opposite lit lower buttons), then push lower buttons (opposite lit upper buttons).

The oscillating strategy on other Petersen graphs

Even when $n \neq 2k$, the oscillating strategy, repeated a finite number of times, will suffice to solve all solvable games on certain Petersen graphs. The proof for each such result is much like the proof of Theorem 1, but matters can be a bit more subtle since the oscillating matrix need not be zero; it need only be the case that every solvable game is in its nullspace.

Table 2 shows the regular Petersen graphs with $n \leq 72$ (and $k < n/2$) for which an oscillating strategy suffices to solve every solvable game. A minimal oscillation sequence is given for each such graph. Playable versions of these games can be found online at [Antonick 2013; Torrence 2016].

References

[Anderson and Feil 1998] M. Anderson and T. Feil, "Turning lights out with linear algebra", *Math. Mag.* **71**:4 (1998), 300–303. MR

[Antonick 2013] G. Antonick, "Numberplay: Robert Torrence's Lights Out", *New York Times* (September 16, 2013).

[Barua and Ramakrishnan 1996] R. Barua and S. Ramakrishnan, "σ-game, σ^+-game and two-dimensional additive cellular automata", *Theoret. Comput. Sci.* **154**:2 (1996), 349–366. MR Zbl

[Goldwasser et al. 1997] J. Goldwasser, W. Klostermeyer, and G. Trapp, "Characterizing switch-setting problems", *Linear and Multilinear Algebra* **43**:1–3 (1997), 121–135. MR Zbl

[Scherphuis 2012] J. Scherphuis, "Lights Out on a graph", 2012, http://www.jaapsch.net/puzzles/lograph.htm.

[Sutner 1988] K. Sutner, "On σ-automata", *Complex Systems* **2**:1 (1988), 1–28. MR Zbl

[Sutner 1990] K. Sutner, "The σ-game and cellular automata", *Amer. Math. Monthly* **97**:1 (1990), 24–34. MR Zbl

[Sutner 2000] K. Sutner, "σ-automata and Chebyshev-polynomials", *Theoret. Comput. Sci.* **230**:1–2 (2000), 49–73. MR Zbl

[Torrence 2016] B. Torrence, "Lights Out on Petersen graphs", 2016, http://faculty.rmc.edu/btorrenc/LightsOut.

Received: 2013-08-30 Revised: 2015-05-25 Accepted: 2015-11-06

btorrenc@rmc.edu *Department of Mathematics, Randolph–Macon College, P.O. Box 5005, Ashland, VA 23005, United States*

rtorrenc@vt.edu *Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.733

A family of elliptic curves of rank ≥ 4

Farzali Izadi and Kamran Nabardi

(Communicated by Ken Ono)

In this paper we consider a family of elliptic curves of the form $y^2 = x^3 - c^2 x + a^2$, where (a, b, c) is a primitive Pythagorean triple. First we show that the rank is positive. Then we construct a subfamily with rank ≥ 4.

1. Introduction

As is well known, an elliptic curve E over a field $\mathbb{K}$ can be explicitly expressed by the generalized Weierstrass equation of the form

$$E : y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6,$$

where $a_1, a_2, a_3, a_4, a_6 \in \mathbb{K}$. In this paper we are interested in the case where $\mathbb{K} = \mathbb{Q}$. By the Mordell–Weil theorem [Washington 2008], every elliptic curve over $\mathbb{Q}$ has a commutative group $E(\mathbb{Q})$ which is finitely generated, i.e., $E(\mathbb{Q}) \cong \mathbb{Z}^r \times E(\mathbb{Q})_{\text{tors}}$, where r is a nonnegative integer and $E(\mathbb{Q})_{\text{tors}}$ is the subgroup of elements of finite order in $E(\mathbb{Q})$. This subgroup is called the torsion subgroup of $E(\mathbb{Q})$ and the integer r is called the rank of E and is denoted by rank E.

By Mazur's theorem [Silverman and Tate 1992], the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ is one of the following 15 groups: $\mathbb{Z}/n\mathbb{Z}$ with $1 \leq n \leq 10$ or $n = 12$ or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2m\mathbb{Z}$ with $1 \leq m \leq 4$. Besides, it is not known which values of rank r are possible. The folklore conjecture is that a rank can be arbitrarily large, but it seems to be very hard to find examples with large ranks. The current record is an example of an elliptic curve over $\mathbb{Q}$ with rank ≥ 28, found by Elkies in May 2006 (see [Dujella 2012]). Having classified the torsion part, one is interested in seeing whether or not the rank is unbounded among all the elliptic curves. There is no known guaranteed algorithm to determine the rank and it is not known which integers can occur as ranks.

Specialization is a significant technique for finding a lower bound for the rank of a family of elliptic curves. One can consider an elliptic curve on the rational function field $\mathbb{Q}(T)$ and then obtain elliptic curves over $\mathbb{Q}$ by specializing the variable T to suitable values $t \in \mathbb{Q}$ (see [Silverman 1994, Chapter III, Theorem 11.4] for more

MSC2010: primary 11G05; secondary 14H52, 14G05.
Keywords: elliptic curves, rank, Pythagorean triple.

details). Using this technique, Nagao and Kouya [1994] found curves of rank ≥ 21, and Fermigier [1996] obtained a curve of rank ≥ 22.

In order to determine r, one should find the generators of the free part of the Mordell–Weil group. Determining the *associated height matrix* is a useful technique for finding a set of generators. In the following, we briefly describe it.

Let $m/n \in \mathbb{Q}$, where $\gcd(m, n) = 1$. Then the *height* of m/n is defined by

$$h\left(\frac{m}{n}\right) = \log\bigl(\max\{|m|, |n|\}\bigr).$$

Corresponding to $P = (x, y) \in E(\mathbb{Q})$, we define

$$H(P) = h(x) \quad \text{and} \quad \hat{h}(P) = \frac{1}{2}\lim_{N\to\infty} \frac{H(2^N \cdot P)}{4^N},$$

where $H(P)$ is called the *canonical height* of $P \in E(\mathbb{Q})$. The Néron–Tate pairing to an elliptic curve is defined by

$$\langle \cdot, \cdot \rangle : E(\mathbb{Q}) \times E(\mathbb{Q}) \to \mathbb{R}, \quad \langle P, Q\rangle = \hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q).$$

The *associated height matrix* to $\{P_i\}_{i=1}^r$ is

$$\mathcal{H} := \bigl(\langle P_i, P_j\rangle\bigr)_{1\leq i\leq r,\, 1\leq j\leq r}.$$

If $\det \mathcal{H} \neq 0$, then the points $\{P_i\}_{i=1}^r$ are linearly independent and $\operatorname{rank} E(\mathbb{Q}) \geq r$ (see[Silverman 1994, Chapter III] for more details and proofs).

In this work we deal with a family of elliptic curves which are related to the Pythagorean triples and, by using both the specialization and, the associated height matrix techniques, prove the following theorem.

Main Theorem 1.1. *Let (a, b, c) be a primitive Pythagorean triple. Then, there are infinitely many elliptic curves of the form*

$$E : y^2 = x^3 - c^2 x + a^2 \tag{1-1}$$

with rank ≥ 4.

If (a, b, c) is a primitive Pythagorean triple, then one can easily check that $a = i^2 - j^2$, $b = 2ij$, and $c = i^2 + j^2$, where $\gcd(i, j) = 1$, and i, j have opposite parity. So, we can consider (1-1) as

$$E_{i,j} : y^2 = x^3 - (i^2 + j^2)^2 x + (i^2 - j^2)^2. \tag{1-2}$$

It is clear that two points $P_{i,j} = (0,\ i^2 - j^2)$ and $Q_{i,j} = (i^2 + j^2,\ i^2 - j^2)$ are on (1-2) and so $\operatorname{rank} E_{i,j} > 0$. In the next section, we construct a subfamily with rank ≥ 3.

2. A subfamily with rank ≥ 3

First, we look at (1-2) as a one-parameter family by letting

$$a = t^2 - 1, \quad b = 2t, \quad c = t^2 + 1, \tag{2-1}$$

where $t \in \mathbb{Q}$. Then, instead of (1-2) one can take

$$E_t : y^2 = x^3 - (t^2+1)^2 x + (t^2-1)^2, \quad t \in \mathbb{Q}. \tag{2-2}$$

Lemma 2.1. *There are infinitely many elliptic curves of the form (2-2) with rank* ≥ 3.

Proof. Clearly we have two points

$$P_t = (0,\ t^2-1), \quad Q_t = (t^2+1,\ t^2-1). \tag{2-3}$$

We impose another point in (2-2) with x-coordinate 1. This implies $1-4t^2$ is a square, say v^2. Then $1-4t^2 = v^2$ defines a circle of the form $(2t)^2 + v^2 = 1$. Hence

$$t = \frac{\alpha}{\alpha^2+1}, \quad v = \frac{\alpha^2-1}{\alpha^2+1}, \tag{2-4}$$

with $\alpha \in \mathbb{Q}$. Then, instead of (2-2), one can take

$$E_\alpha : y^2 = x^3 - \left(\left(\frac{\alpha}{\alpha^2+1}\right)^2 + 1\right)^2 x + \left(\left(\frac{\alpha}{\alpha^2+1}\right)^2 - 1\right)^2, \tag{2-5}$$

having three points

$$P_\alpha = \left(0,\ \left(\frac{\alpha}{\alpha^2+1}\right)^2 - 1\right), \quad Q_\alpha = \left(\left(\frac{\alpha}{\alpha^2+1}\right)^2 + 1,\ \left(\frac{\alpha}{\alpha^2+1}\right)^2 - 1\right), \quad R_\alpha = \left(1,\ \frac{\alpha^2-1}{\alpha^2+1}\right).$$

When we specialize to $\alpha = 2$, we obtain a set of points $S = \{P_2, Q_2, R_2\} = \left\{\left(0, \frac{-21}{25}\right), \left(\frac{29}{25}, \frac{-21}{25}\right), \left(1, \frac{3}{5}\right)\right\}$ on

$$E_2 : y^2 = x^3 - \tfrac{841}{25} x + \tfrac{44}{25}. \tag{2-6}$$

Using SAGE, one can easily check that the associated height matrix of S has nonzero determinant $\approx 22.879895 \neq 0$ showing that these three points are independent and so rank $E_2 \geq 3$. The specialization result of Silverman [1994] implies that for all but finitely many rational numbers, rank $E_\alpha \geq 3$. □

3. Proof of the main theorem

We impose another point with x-coordinate $-2\alpha/(\alpha^2+1)$ in (2-5). Hence we want $1 + 2\alpha/(\alpha^2+1)$ to be a square. It suffices that α^2+1 is a square, say β^2. Therefore,

$$\alpha = \frac{2m}{1-m^2}, \quad \beta = \frac{m^2+1}{1-m^2}, \tag{3-1}$$

where $m \in \mathbb{Q}$. From the above expressions, one can transform (2-5) to

$$E_m : y^2 = x^3 - \frac{(m^8+8m^6-2m^4+8m^2+1)^2}{(2m^2+m^4+1)^4} x + \frac{(m^8+14m^4+1)^2}{(2m^2+m^4+1)^4}. \tag{3-2}$$

So we get the four points

$$P_m = (0, \gamma), \qquad Q_m = \left(\frac{m^8+8m^6-2m^4+8m^2+1}{(m^2+1)^4}, \gamma \right),$$
$$R_m = \left(1, \frac{(m^2-2m-1)(m^2+2m-1)}{(m^2+1)^2}\right), \qquad S_m = \left(\frac{4m(m^2-1)}{m^4+2m^2+1}, \frac{(m^2-2m-1)}{m^2+1}\gamma \right),$$

where

$$\gamma = \frac{(m^4-2m^3+2m^2+2m+1)(m^4+2m^3+2m^2-2m+1)}{(m^2+1)^4}.$$

By specialization to $m = 2$ in (3-2), we have

$$E_2 : y^2 = x^3 - \tfrac{591361}{390625}x + \tfrac{231361}{390625}, \tag{3-3}$$

and the set of points $S = \{P_2, Q_2, R_2, S_2\} = \left\{ \left(0, \frac{481}{625}\right), \left(\frac{769}{625}, \frac{481}{625}\right), \left(1, \frac{7}{5}\right), \left(\frac{24}{25}, \frac{481}{3125}\right) \right\}$ on it. The associated height matrix of these four points has nonzero determinant $\approx 722.7181 \neq 0$ showing that these points are independent and so rank $E_2 \geq 4$. However, by using SAGE we see that rank $E_2 = 5$. Again, by specialization, we can conclude that for all but finitely many elliptic curves of the form (3-2), we have rank ≥ 4.

Acknowledgements

The authors express their hearty thanks to the anonymous referee for a careful reading of the paper and for many comments and remarks which improved its quality.

References

[Dujella 2012] A. Dujella, "High rank elliptic curves with prescribed torsion", 2012, available at https://web.math.pmf.unizg.hr/~duje/tors/tors.html.

[Fermigier 1996] S. Fermigier, "Construction of high-rank elliptic curves over $\mathbb{Q}$ and $\mathbb{Q}(t)$ with non-trivial 2-torsion (extended abstract)", pp. 115–120 in *Algorithmic number theory* (Talence, 1996), edited by H. Cohen, Lecture Notes in Computer Science **1122**, Springer, Berlin, 1996. MR 1446503 Zbl 0890.11020

[Nagao and Kouya 1994] K.-I. Nagao and T. Kouya, "An example of elliptic curve over $\mathbb{Q}$ with rank ≥ 21", *Proc. Japan Acad. Ser. A Math. Sci.* **70**:4 (1994), 104–105. MR 1276883 Zbl 0832.14022

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 1312368 Zbl 0911.14015

[Silverman and Tate 1992] J. H. Silverman and J. Tate, *Rational points on elliptic curves*, Springer, New York, NY, 1992. MR 1171452 Zbl 0752.14034

[Washington 2008] L. C. Washington, *Elliptic curves: number theory and cryptography*, 2nd ed., Chapman and Hall, Boca Raton, FL, 2008. MR 2404461 Zbl 1200.11043

Received: 2014-05-07 Revised: 2015-09-30 Accepted: 2015-10-01

f.izadi@urmia.ac.ir *Department of Mathematics, Urmia University, Urmia, Iran*

nabardi@azaruniv.edu *Department of Mathematics, Azarbaijan Shahid Madani University, Tabriz, Iran*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.737

Splitting techniques and Betti numbers of secant powers

Reza Akhtar, Brittany Burns, Haley Dohrmann,
Hannah Hoganson, Ola Sobieska and Zerotti Woods

(Communicated by Scott T. Chapman)

Using ideal-splitting techniques, we prove a recursive formula relating the Betti numbers of the secant powers of the edge ideal of a graph H to those of the join of H with a finite independent set. We apply this result in conjunction with other splitting techniques to compute these Betti numbers for wheels, complete graphs and complete multipartite graphs, recovering and extending some known results about edge ideals.

1. Introduction

Let R be a polynomial ring in finitely many variables over a base field $\mathbb{K}$. One approach to studying modules over R is by constructing free resolutions and studying properties of these. If M is a finitely generated graded R-module, Hilbert's syzygy theorem implies that there exists a free resolution with only finitely many terms. Furthermore, one can show that among these free resolutions, there is one which is minimal (in a sense which will be made precise later), and thereby defines a collection of integers, the *Betti numbers* of M. Of particular interest is the case when the module in question is an ideal of R. Even more specifically, if G is a simple graph with vertices $v_1, \dots, v_n$, its *edge ideal*, $I(G)$, is the ideal in $R = \mathbb{K}[x_1, \dots, x_n]$ generated by the monomials $x_i x_j$ such that $v_i v_j$ is an edge of G. The edge ideal was first defined by Villarreal [1995] and has attracted considerable interest as an algebraic object which encodes combinatorial information. In recent years, much attention has been devoted to studying the Betti numbers of edge ideals; see, for example, [Emtander 2009; Francisco et al. 2009; Hà and Van Tuyl 2007; 2008; Jacques 2004]. Betti numbers are also of interest in algebraic geometry [Sidman and Vermeire 2009; 2011], as the edge ideal defines a (not necessarily irreducible) variety in n-dimensional projective space over $\mathbb{K}$.

MSC2010: primary 13D02; secondary 05C25.
Keywords: Betti number, edge ideal, secant power, complete graph, complete bipartite graph.

A more general problem is that of computing the Betti numbers of the *secant powers* of the edge ideal. The actual definition of the secant powers of an ideal is somewhat delicate, but the idea is not hard to grasp. The first secant power is the ideal itself, and if V is the variety in n-dimensional projective space over $\mathbb{K}$ defined by the ideal, then its r-th secant power is the ideal which defines the r-th secant variety of V. The Betti numbers of secant powers of the edge ideal have also been studied in the literature (see [Cranfill 2009; Rosen 2009], and especially [Sidman and Vermeire 2009; 2011]) but not nearly as extensively as those of the edge ideal. For convenience of reference, we will use the phrase "Betti numbers of G" throughout this article as shorthand for "Betti numbers of the secant powers of the edge ideal of G".

In his Ph.D. thesis, Jacques [2004] studied and computed the Betti numbers of the edge ideals corresponding to various classes of graphs, including cycles, paths, forests, complete graphs, and complete bipartite graphs. His main tool was a formula of Hochster [1977] which expresses the Betti numbers of a Stanley–Reisner ring over a simplicial complex in terms of the (simplicial) homology of the complex. Using this formula, Jacques was able to give exact computations of all the Betti numbers of complete graphs and complete bipartite graphs. His techniques have been applied in several works since (for example, [Emtander 2009]) and have proven to be quite fruitful.

An alternative approach to computing Betti numbers of edge ideals was initiated by Tài Hà and Van Tuyl [2007; 2008]; see in particular Theorems 3.6 and 4.6 of their 2007 paper. This technique, called *ideal splitting*, goes back to the work of Eliahou and Kervaire [1990] in the ungraded case and Fatabbi [2001] in the graded case. The idea is to decompose the (monomial) ideal under consideration into simpler pieces, and make use of a formula relating the Betti numbers of the pieces to the Betti numbers of the original ideal. The advantage of this approach is that it obviates the need to compute simplicial homology groups and allows, at least in some cases, for the calculation of Betti numbers by induction.

The present article is written in the spirit of [Hà and Van Tuyl 2007], but the notion of ideal splitting is applied in a different way, and in a different setting. Using a combinatorial description of higher secant ideals due to Sturmfels and Sullivant [2006], we derive a recursive formula (Theorem 4.4) which allows us to relate the Betti numbers of the join of a graph with a finite independent set to the Betti numbers of the graph itself. Since complete graphs and complete bipartite graphs can both be constructed by iterating this type of join operation, one can use this formula to compute the Betti numbers of *all the secant powers* of their edge ideals. In the process, we recover Jacques's calculations (for the edge ideal itself) by purely combinatorial means, without recourse to Hochster's formula. We emphasize that all our results are independent of the choice of base field $\mathbb{K}$.

2. Preliminaries

We now provide some background on minimal free resolutions; more detail may be found in any standard book on the subject, for example [Eisenbud 1995].

Throughout this article, we fix a base field $\mathbb{K}$. Let $x_1, \dots, x_t$ be independent indeterminates and $R = \mathbb{K}[x_1, \dots, x_t]$. Then R is an $\mathbb{N}$-graded ring in the natural way: $R = \bigoplus_e R_e$, where R_e is the $\mathbb{K}$-vector space spanned by the monomials in $x_1, \dots, x_t$ of total degree e. Note also that R has a unique maximal ideal $\mathfrak{m}$ consisting of all elements of positive degree. For any integer d, we denote by $R(d)$ the graded ring whose degree-e part is R_{d+e}. An ideal $I \subseteq R$ is called a *monomial ideal* if it is generated by monomials.

Now suppose that I is an ideal of R. Because R/I is finitely generated as an R-module, Hilbert's syzygy theorem [Eisenbud 1995, Corollary 19.7] implies that it has a finite resolution by free modules; that is, there exists an integer $n \le t+1$, finitely generated free R-modules $F_0, \dots, F_n$, and R-module homomorphisms $\phi_i : F_i \to F_{i-1}$, for $i = 1, \dots, n$, and $\phi_0 : F_0 \to R/I$ such that

$$0 \to F_n \xrightarrow{\phi_n} F_{n-1} \xrightarrow{\phi_{n-1}} \cdots \to F_1 \xrightarrow{\phi_1} F_0 \xrightarrow{\phi_0} R/I \to 0$$

is an exact sequence.

It can be shown [Eisenbud 1995, Theorem 20.2] that R/I has a *minimal* free resolution of the above form, meaning that $\phi_i(F_i) \subseteq \mathfrak{m}F_{i-1}$ for $i = 1, \dots, n$. Furthermore, any two minimal free resolutions of I are isomorphic (as chain complexes), so the F_i are uniquely determined (as R-modules) up to isomorphism. Thus, each free module F_i may be written $\bigoplus_j R(-j)^{b_{i,j}(I)}$ in such a way so as to ensure that each of the maps $\phi_1, \dots, \phi_n$ is a homomorphism of *graded* R-modules. Note that since F_i is finitely generated as an R-module, $b_{i,j}(I) = 0$ for all but finitely many j. The numbers $b_{i,j}(I)$ are called the (graded) *Betti numbers* of I. It is clear that for any R and nonzero ideal $I \subseteq R$, we have $b_{0,0}(I) = 1$ and $b_{0,j}(I) = 0$ for $j \neq 0$.

Since exactness is preserved under flat base change, we immediately have:

Proposition 2.1. *If R' is a flat graded R-algebra, then for any ideal I,*

$$b_{i,j}(I \otimes_R R') = b_{i,j}(I).$$

We are interested in the case $R = \mathbb{K}[x_1, \dots, x_m]$, $R' = \mathbb{K}[x_1, \dots, x_m, y_1, \dots, y_n]$, where $x_1, \dots, x_n, y_1, \dots, y_m$ are independent indeterminates. In this situation, $I \otimes_R R'$ is simply the extension of the ideal $I \subseteq R$ to the larger ring R'.

It is also worth recording a standard result which follows directly from the construction of the Koszul complex:

Proposition 2.2 [Eisenbud 1995, Corollary 19.3]. *For $i \ge 0$, we have $b_{i,i}(\mathfrak{m}) = \binom{t}{i}$.*

We will also be studying the secant powers of various monomial ideals in R. Since the definition itself is rather complicated and formulated in greater generality than we will need, we omit it here and instead refer the interested reader to [Simis and Ulrich 2000] or [Sturmfels and Sullivant 2006] for details. The points we will need may be summarized as follows. There is an operation $*$ on ideals of R called the *join*, which is both associative and commutative. If I is a ideal of R, we define its *secant powers* by $I^{\{0\}} = \mathfrak{m}$, $I^{\{1\}} = I$, and, for $r > 1$, $I^{\{r\}} = I * I^{\{r-1\}}$. Moreover, if I is a monomial ideal, then there is a convenient method for computing the generators of its secant powers in terms of its own generators (see [Simis and Ulrich 2000, Proposition 3.1] for details). The "secant" terminology comes from algebraic geometry: if one considers I as defining a variety V in n-dimensional projective space over $\mathbb{K}$, then $I^{\{r\}}$ defines the r-fold secant variety of V.

3. Edge ideals and splitting

In this section, we define the edge ideal of a graph and recall a result which allows for a simple combinatorial description of a minimal generating set for each of its secant powers. Throughout this article, all graphs are assumed to be simple, with a finite vertex set. Given a subset S of vertices in a graph G, we denote by G_S the subgraph of G induced by S, i.e., the graph whose vertex set is S and whose edge set consists of those edges of G, both of whose endpoints lie in S. We denote by K_m the complete graph on m vertices and by $\overline{G}$ the complement of a graph G. If G and H are graphs with disjoint vertex sets, the *join* of G and H, denoted $G \vee H$, is the graph whose vertex set is $V(G) \cup V(H)$, and whose edge set is $E(G) \cup E(H) \cup \{uv : u \in V(G),\ v \in V(H)\}$. (This join operation on graphs is not related to the join of ideals defined in Section 2.) Intuitively, one may think of the join of two graphs as constructed by taking disjoint copies of each and adding all possible edges with one endpoint in each of the two graphs. The join operation on graphs is easily seen to be associative. Finally, we denote by $\chi(G)$ the chromatic number of G; this is the smallest positive integer k such that there exists an assignment of an integer from $\{1, \ldots, k\}$ to each vertex of G in such a way that no two adjacent vertices are labeled with the same integer. For further details on graph theory, we refer the reader to [West 1996] or any other standard textbook on the subject.

Let G be graph with vertex set $V(G) = \{v_1, \ldots, v_n\}$. Let $x_1, \ldots, x_n$ be independent indeterminates, and let $I(G)$ be the ideal of $R = \mathbb{K}[x_1, \ldots, x_n]$ generated by all monomials $x_i x_j$ such that $v_i v_j$ is an edge of G; we call $I(G)$ the *edge ideal* of G. If $S = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$, we denote by M_S the monomial $x_{i_1} \cdots x_{i_m} \in \mathbb{K}[x_1, \ldots, x_n]$. We also define

$$\mathcal{C}_r(G) = \big\{S \subseteq V(G) : \chi(G_S) = r + 1 \text{ and } \chi(G_T) \leq r \text{ for all proper } T \subseteq S\big\}.$$

Sturmfels and Sullivant have given a convenient combinatorial description of the secant ideals $I(G)^{\{r\}}$.

Theorem 3.1 [Sturmfels and Sullivant 2006, Theorem 3.2]. *The ideal $I(G)^{\{r\}}$ is generated by $\{M_S : S \subseteq V(G)$ and $\chi(G_S) \geq r+1\}$. A minimal generating set for $I(G)^{\{r\}}$ is given by $\mathcal{S}_r(G) = \{M_S : S \in \mathcal{C}_r(G)\}$.*

The following elementary fact about monomial ideals is well known:

Proposition 3.2. *Suppose I and J are monomial ideals in a polynomial ring R over a field, generated (respectively) by monomial sets A and B. Then $I \cap J$ is also a monomial ideal in R and is generated by $\{\mathrm{lcm}(a, b) : a \in A,\ b \in B\}$.*

We now define the notion of a *splittable* ideal, due to Eliahou and Kervaire.

Definition 3.3 [Eliahou and Kervaire 1990]. A monomial ideal I in a polynomial ring R (over a field) is called *splittable* if there exist ideals J and K of R and minimal generating sets $\mathcal{G}(I)$, $\mathcal{G}(J)$, and $\mathcal{G}(K)$ for I, J, and K (respectively), and a generating set $\mathcal{G}(J \cap K)$ for $J \cap K$ such that:

(1) $I = J + K$.

(2) $\mathcal{G}(I)$ is the disjoint union of $\mathcal{G}(J)$ and $\mathcal{G}(K)$.

(3) There are functions $\phi : \mathcal{G}(J \cap K) \to \mathcal{G}(J)$ and $\psi : \mathcal{G}(J \cap K) \to \mathcal{G}(K)$ such that:
   (a) For all $u \in \mathcal{G}(J \cap K)$, we have $u = \mathrm{lcm}(\phi(u), \psi(u))$.
   (b) For every subset $C \subseteq \mathcal{G}(J \cap K)$, both $\mathrm{lcm}(\phi(C))$ and $\mathrm{lcm}(\psi(C))$ strictly divide $\mathrm{lcm}(C)$.

In this situation, we say that $I = J + K$ is a *splitting* of I and refer to the pair (ϕ, ψ) as a *splitting function*.

Remark. In the original formulation of this definition, the generating set for $J \cap K$ was also required to be minimal. However, since every generating set contains a minimal generating set, the two formulations are in fact equivalent.

The following result of Fatabbi relates splittability to the computation of the Betti numbers of the ideal in question.

Theorem 3.4 [Fatabbi 2001, Proposition 3.2]. *Suppose I is a splittable monomial ideal in a polynomial ring over a field, with splitting $I = J + K$. Then*

$$b_{i,j}(I) = b_{i,j}(J) + b_{i,j}(K) + b_{i-1,j}(J \cap K)$$

for all integers $i \geq 1$ and j, provided we interpret $b_{0,j}(J \cap K)$ as 0.

4. Main result

The goal of this section is to develop a formula relating the Betti numbers of the join of a graph H with an edgeless graph to those of H itself.

Let $v_1, \ldots, v_n$ be an ordering of the vertices in a graph H. Now let $w_1, \ldots, w_m$ be new vertices and, for $1 \leq \ell \leq m$, define H_ℓ as the join of H with the edgeless graph on $W = \{w_1, \dots, w_\ell\}$. If we set $H_0 = H$, then we may view each H_ℓ, for $0 \le \ell \le m$, as isomorphic to $H \vee \overline{K_\ell}$. Now define $R = R_0 = \mathbb{K}[x_1, \dots, x_n]$ and $R_\ell = \mathbb{K}[x_1, \dots, x_n, y_1, \dots, y_\ell]$ for $1 \le \ell \le m$.

Lemma 4.1. *Suppose $1 \le \ell \le m$. Then the elements of $\mathcal{C}_r(H_\ell)$ are of two types*:

(i) *subsets $S \subseteq V(H)$ such that $S \in \mathcal{C}_r(H)$,*

(ii) *subsets of the form $S' \cup \{w\}$, where $S' \in \mathcal{C}_{r-1}(H)$ and $w \in W$.*

Proof. For convenience, set $H' = H_\ell$, and suppose $S \in \mathcal{C}_r(H')$. If $S \subseteq V(H)$, then clearly $S \in \mathcal{C}_r(H)$, so suppose S is not contained in $V(H)$. We claim that S contains exactly one of $w_1, \dots, w_\ell$. Suppose to the contrary that w_i and w_j are both in S, where $1 \le i < j \le \ell$, and let $T = S - \{w_j\}$. Let $f' : T \to \{1, \dots, t\}$ be a proper coloring of T. Since w_i is adjacent to all vertices of $T \cap V(H)$, we have $f'(w_i) \neq f'(v)$ for all $v \in T$. Extend f' to a function $f : S \to \{1, \dots, t\}$ by setting $f(w_j) = f'(w_i)$. Since w_j is not adjacent to any vertex of $S \cap W$ but is adjacent to all vertices in $T \cap V(H)$, f is a proper t-coloring of S. This shows that $\chi(H'_S) \le \chi(H'_T)$. Since obviously $\chi(H'_T) \le \chi(H'_S)$, it follows that $\chi(H'_T) = \chi(H'_S)$, contradicting the hypothesis $S \in \mathcal{C}_r(H')$. □

We refer to members of $\mathcal{C}_r(H_\ell)$ as either of type (i) or type (ii), according to their classification in Lemma 4.1.

Define $A_{0,0} = I(H)^{\{r\}}$, and, for $1 \le k \le \ell \le m$, let $A_{k,\ell}$ be the ideal of R_ℓ generated by all M_S such that $S \in \mathcal{C}_r(H_\ell)$ is of type (ii) and $W \cap S = \{w_\ell\}$. Also define $B_{0,0} = 0$ and $B_{k,\ell} = \sum_{j=1}^{k} A_{j,\ell}$ for $1 \le k \le \ell \le m$. Note further that if $0 \le k \le \ell \le \ell' \le m$, then by construction $A_{k,\ell'} = A_{k,\ell} \otimes_{R_\ell} R_{\ell'}$ and $B_{k,\ell'} = B_{k,\ell} \otimes_{R_\ell} R_{\ell'}$, so Proposition 2.1 implies

$$b_{i,j}(A_{k,\ell}) = b_{i,j}(A_{k,\ell'}) \quad \text{and} \quad b_{i,j}(B_{k,\ell}) = b_{i,j}(B_{k,\ell'}). \tag{1}$$

Lemma 4.2. *For $1 \le k \le \ell \le m$, there are isomorphisms*

$$A_{k,\ell} \cong [I(H)^{\{r-1\}} \otimes_R R_\ell](-1) \quad \textit{and} \quad A_{k,\ell} \cap I(H)^{\{r\}} \cong [I(H)^{\{r\}} \otimes_R R_\ell](-1)$$

of graded R'-modules, and thus

$$b_{i,j}(A_{k,\ell}) = b_{i,j-1}(I(H)^{\{r-1\}}), \quad b_{i,j}(A_{k,\ell} \cap I(H)^{\{r\}}) = b_{i,j-1}(I(H)^{\{r\}}).$$

Proof. By Lemma 4.1, $A_{k,\ell}$ is generated by monomials of the form $y_k M_{S'}$, where $S' \in \mathcal{C}_{r-1}(H)$. Thus, $A_{k,\ell} = y_k(I(H)^{\{r-1\}} \otimes_R R')$, which is isomorphic (as a graded R'-module) to $[I(H)^{\{r-1\}} \otimes_R R'](-1)$. By Proposition 2.1,

$$b_{i,j}(A_{k,\ell}) = b_{i,j-1}(I(H)^{\{r-1\}}),$$

as predicted by the formula. Likewise, by Proposition 3.2, we see that $A_{k,\ell} \cap I(H)^{\{r\}}$ is generated by monomials of the form $y_k M_{S'}$, where $S' \in \mathcal{C}_r(H)$. Arguing as above, we have $A_{k,\ell} \cap I(H)^{\{r\}} \cong I(H)^{\{r\}}(-1)$, whence the result. □

Lemma 4.3. *Let $r \geq 1$ and $1 \leq k \leq \ell \leq m$. Then there are splittings*

$$B_{k,\ell} = B_{k-1,\ell} + A_{k,\ell}, \quad B_{k,\ell} \cap I(H)^{\{r\}} = B_{k-1,\ell} \cap I(H)^{\{r\}} + A_{k,\ell} \cap I(H)^{\{r\}}.$$

Thus,

$$b_{i,j}(B_{k,\ell}) = b_{i,j}(B_{k-1,\ell}) + b_{i,j-1}(I(H)^{\{r-1\}}) + b_{i-1,j-1}(B_{k-1,\ell}),$$

$$b_{i,j}(B_{k,\ell} \cap I(H)^{\{r\}}) = b_{i,j}(B_{k-1,\ell} \cap I(H)^{\{r\}}) + b_{i,j-1}(I(H)^{\{r\}}) + b_{i-1,j-1}(B_{k-1,\ell} \cap I(H)^{\{r\}}).$$

Proof. We will prove the first formula, the second being similar, mutatis mutandis. By Lemma 4.1, a set of minimal generators for $A_{k,\ell}$ is given by $y_k M_{S'}$, where $S' \in \mathcal{C}_{r-1}(H)$. By Proposition 3.2, a generating set for $B_{k-1,\ell} \cap A_{k,\ell}$ is given by the set of monomials $y_k M_{S'}$, where $S' \in \mathcal{C}_r(H_{k-1})$. Now let $\mu(S') = \max\{t : v_t \in S'\}$ and choose $T(S') \subseteq S' - \{v_{\mu(S')}\}$ such that $T(S') \in \mathcal{C}_{r-1}(H_{k-1})$. Observe also that $B_{k-1,\ell} \cap A_{k,\ell} \cong B_{k-1,\ell}(-1)$.

We claim that the correspondence $y_k M_{S'} \mapsto (M_{S'}, y_k M_{T(S')})$ defines a splitting function. The first and second conditions of Definition 3.3 are clearly satisfied. For the last condition, let $C = \{y_k M_{S'_d} : d \in D\}$ (where D is some set indexing the monomials) be a subset of the generating set for $B_{k-1,\ell} \cap A_{k,\ell}$ described above. Then the first coordinate of the image of any element of C under the above function does not involve the variable y_k. Furthermore, the second coordinate does not involve the variable x_M, where $M = \max_{d \in D} \mu_{(S'_d)}$. This shows that $B_{k,\ell} = B_{k-1,\ell} + A_{k,\ell}$ defines a splitting. The remaining formulas follow from Theorem 3.4. □

We now come to our main result.

Theorem 4.4. *Let H be graph and r, m positive integers. Then for all j,*

$$b_{1,j}(I(H_m)^{\{r\}}) = b_{1,j}(I(H)^{\{r\}}) + m b_{1,j-1}(I(H)^{\{r-1\}}),$$

and for $i \geq 2$,

$$b_{i,j}(I(H_m)^{\{r\}}) = b_{i,j}(I(H_{m-1})^{\{r\}}) + b_{i,j-1}(I(H)^{\{r-1\}}) + b_{i-1,j-1}(I(H_{m-1})^{\{r\}}).$$

Proof. Let $1 \leq \ell \leq m$. The generators of $I(H_\ell)^{\{r\}}$ are described by Lemma 4.1: $J = I(H)^{\{r\}}$ is the ideal of R' generated by the monomials M_S, where S is of type (i), and $K = B_{\ell,\ell}$ is the ideal generated by M_S for S of type (ii). We claim that

$$I(H_\ell)^{\{r\}} = I(H)^{\{r\}} + B_{\ell,\ell} \tag{2}$$

is in fact a splitting.

It is clear from the above description of the generators of $I(H_\ell)^{\{r\}}$ that the second condition of Definition 3.3 is satisfied, so it remains to construct a splitting function. By Proposition 3.2 and Lemma 4.1, a generator $M_S \in \mathcal{G}(J \cap K)$ is a monomial of the form $y_j M_{S'}$, where $1 \leq j \leq \ell$ and $S' \in \mathcal{C}_r(H)$. Let $\mu(S') = \max\{i : v_i \in S'\}$;

then choose $T(S') \subseteq S' - \{v_{\mu(S')}\}$ such that $T(S') \in \mathcal{C}_{r-1}(H)$. We claim that $y_j M_{S'} \mapsto (M_{S'}, y_j M_{T(S')})$ defines a splitting function.

As before, the first and second conditions of Definition 3.3 are clearly satisfied. With notation as above, let $C = \{y_{j_d} M_{S'_d} : d \in D, 1 \leq j_a \leq \ell\}$ be a subset of the generators of $J \cap K$. Now the monomial $\operatorname{lcm}(C)$ involves some indeterminate from among $y_1, \ldots, y_\ell$; however, the first coordinate of its image under the proposed function does not involve any of the y_j. Furthermore, the second coordinate does not involve x_N, where $N = \max_{d \in D} \max\{i : v_i \in T(S'_d)\}$, whereas $\operatorname{lcm}(C)$ does. Thus, (2) is a splitting, as claimed.

Applying Theorem 3.4 to (2) implies

$$b_{1,j}(I(H_m)^{\{r\}}) = b_{1,j}(I(H)^{\{r\}}) + b_{1,j}(B_{m,m}).$$

By Lemma 4.3 and (1),

$$b_{1,j}(B_{m,m}) = b_{1,j}(B_{m-1,m-1}) + b_{1,j-1}(I(H)^{\{r-1\}}).$$

Applying this successively yields

$$\begin{aligned} b_{1,j}(B_{m,m}) &= b_{1,j}(B_{0,0}) + m b_{1,j-1}(I(H)^{\{r-1\}}) \\ &= b_{1,j}(I(H)^{\{r\}}) + m b_{1,j-1}(I(H)^{\{r-1\}}), \end{aligned}$$

which establishes the first formula.

Now suppose $i \geq 2$. Applying Theorem 3.4 to (2) with $\ell = m$ yields

$$b_{i,j}(I(H_m)^{\{r\}}) = b_{i,j}(I(H)^{\{r\}}) + b_{i,j}(B_{m,m}) + b_{i-1,j}(B_{m,m} \cap I(H)^{\{r\}}),$$

which by Lemma 4.3 may be rewritten as

$$\begin{aligned} b_{i,j}(I(H_m)^{\{r\}}) = b_{i,j}(I(H)^{\{r\}}) &+ b_{i,j}(B_{m-1,m}) + b_{i,j-1}(I(H)^{\{r-1\}}) \\ &+ b_{i-1,j-1}(B_{m-1,m}) \\ &\quad + b_{i-1,j}(B_{m-1,m} \cap I(H)^{\{r\}}) + b_{i-1,j-1}(I(H)^{\{r\}}) \\ &\qquad + b_{i-2,j-1}(B_{m-1,m} \cap I(H)^{\{r\}}). \end{aligned}$$

Applying (1), this becomes

$$\begin{aligned} b_{i,j}(I(H_m)^{\{r\}}) = b_{i,j}(I(H)^{\{r\}}) &+ b_{i,j}(B_{m-1,m-1}) + b_{i,j-1}(I(H)^{\{r-1\}}) \\ &+ b_{i-1,j-1}(B_{m-1,m-1}) + b_{i-1,j}(B_{m-1,m-1} \cap I(H)^{\{r\}}) \\ &+ b_{i-1,j-1}(I(H)^{\{r\}}) + b_{i-2,j-1}(B_{m-1,m-1} \cap I(H)^{\{r\}}). \end{aligned} \tag{3}$$

However, Theorem 3.4 applied to (2) with $\ell = m - 1$ yields

$$b_{i,j}(I(H_{m-1})^{\{r\}}) = b_{i,j}(I(H)^{\{r\}}) + b_{i,j}(B_{m-1,m-1}) + b_{i-1,j}(B_{m-1,m-1} \cap I(H)^{\{r\}}). \tag{4}$$

Subtracting (4) from (3), we obtain

$$\begin{aligned}
b_{i,j}(I(H_m)^{\{r\}}) - b_{i,j}(I(H_{m-1})^{\{r\}})&\\
&= b_{i,j-1}(I(H)^{\{r-1\}}) + b_{i-1,j-1}(B_{m-1,m-1})\\
&\qquad + b_{i-1,j-1}(I(H)^{\{r\}}) + b_{i-2,j-1}(B_{m-1,m-1} \cap I(H)^{\{r\}}).\\
&= b_{i,j-1}(I(H)^{\{r-1\}}) + b_{i-1,j-1}(B_{m-1,m-1} + I(H)^{\{r\}})\\
&= b_{i,j-1}(I(H)^{\{r-1\}}) + b_{i-1,j-1}(I(H_{m-1})^{\{r\}}). \qquad \square
\end{aligned}$$

5. Applications

In this section, we apply Theorem 4.4 to calculate the Betti numbers for some common classes of graphs. To illustrate the key ideas, we begin with the relatively simple case of wheels, and then proceed to the case of complete graphs. Both of these calculations only use the case $m = 1$ of Theorem 4.4 and yield fairly elegant formulas for the Betti numbers. We conclude with the case of complete multipartite graphs, which is technically more complicated. Note that from the discussion of Section 2, we always have $b_{0,0} = 1$ and $b_{0,j} = 0$ for $j \neq 0$; hence we will focus on $b_{i,j}$ when $i \geq 1$.

Wheels. For an integer $n \geq 3$, the *n-cycle*, denoted C_n, is the graph on vertices $v_1, \ldots, v_n$ whose edges are $v_n v_1$ and $v_i v_{i+1}$, where $1 \leq i \leq n-1$. The *n-wheel*, denoted W_n, is the join of C_n with $\overline{K_1}$. To compute the Betti numbers of W_n using Theorem 4.4, we will need the Betti numbers of C_n. For the edge ideal, these were calculated by Jacques [2004, Theorem 7.6.28]: when $j < n$ and $2i \geq j$,

$$b_{i,j}(I(C_n)) = \frac{n}{n-2(j-i)}\binom{j-i}{2i-j}\binom{n-2(j-i)}{j-i}.$$

Moreover, if $n = 3m+1$ or $n = 3m+2$, then $b_{2m+1,n}(I(C_n)) = 1$, and if $n = 3m$, then $b_{2m,n}(I(C_n)) = 2$; all other Betti numbers of $I(C_n)$ are 0. Now if n is even, then $\chi(C_n) = 2$, so $I(C_n)^{\{r\}} = 0$ for $r \geq 2$. If n is odd, then $\chi(C_n) = 3$, so by Theorem 3.1, we have $I(C_n)^{\{2\}}$ is generated by the single monomial $x_1 \cdots x_n$. As such, we have $I(C_n)^{\{2\}} \cong R(-n)$; hence its only nonzero Betti number is $b_{1,n}\big(I(C_n)^{\{2\}}\big) = 1$. Clearly $I(C_n)^{\{r\}} = 0$ for $r \geq 3$.

We now turn to the computation of the Betti numbers of W_n, $n \geq 3$. In the interest of making the presentation more readable, we will express the Betti numbers of W_n in terms of those of C_n and other directly computable quantities. We begin with the edge ideal of W_n.

By Theorem 4.4 and Proposition 2.2,

$$b_{1,j}(I(W_n)) = b_{1,j}(I(C_n)) + b_{1,j-1}(I(C_n)^{\{0\}}) = \begin{cases} 2n & \text{if } j = 2, \\ 0 & \text{if } j \neq 2. \end{cases}$$

If $i \geq 2$, we have $b_{i,j}(I(W_n)) = b_{i,j}(I(C_n)) + b_{i,j-1}(I(C_n)^{\{0\}}) + b_{i-1,j-1}(I(C_n))$, so

$$b_{i.j}(I(W_n)) = \begin{cases} b_{i,i+1}(I(C_n)) + b_{i-1,i}(I(C_n)) + \binom{n}{i} & \text{if } j = i+1, \\ b_{i,j}(I(C_n)) + b_{i-1,j-1}(I(C_n)) & \text{if } j \neq i+1. \end{cases}$$

Turning our attention to the second secant ideal of W_n, we have

$$b_{1,j}(I(W_n)^{\{2\}}) = b_{1,j}(I(C_n)^{\{2\}}) + b_{1,j-1}(I(C_n)),$$

and for $i \geq 2$,

$$b_{i,j}(I(W_n)^{\{2\}}) = b_{i,j}(I(C_n)^{\{2\}}) + b_{i,j-1}(I(C_n)) + b_{i-1,j-1}(I(C_n)^{\{2\}}).$$

Thus, when n is even, $b_{i,j}(I(W_n)^{\{2\}}) = b_{i,j-1}(I(C_n))$ for all $i \geq 1$. When n is odd, we have $b_{i,j}(I(W_n)^{\{2\}}) = b_{i,j-1}(I(C_n)) + \varepsilon_{i,j}$, where $\varepsilon_{1,n} = \varepsilon_{2,n+1} = 1$ and $\varepsilon_{i,j} = 0$ otherwise.

When n is even, $I(W_n)^{\{r\}} = 0$ when $r \geq 3$. Finally, when n is odd, the only subgraph of W_n of chromatic number 4 is W_n itself, so $b_{1,n+1}\big(I(W_n)^{\{3\}}\big) = 1$ is the only nonzero Betti number of $I(W_n)^{\{3\}}$, and of course $I(W_n)^{\{r\}} = 0$ when $r \geq 4$.

Complete graphs. Since $K_n = K_{n-1} \vee \overline{K_1}$, Theorem 4.4 provides a means of calculating its Betti numbers recursively. In fact, there is an elegant formula in closed form which recovers and extends Jacques's computation [2004, Theorem 5.1.1] in the case of the edge ideal.

Theorem 5.1. *Suppose n, i are positive integers and r is a nonnegative integer. Then*

$$b_{i,i+r}(I(K_n)^{\{r\}}) = \binom{i+r-1}{r}\binom{n}{i+r}.$$

If $j \neq i+r$, then $b_{i,j}(I(K_n)^{\{r\}}) = 0$.

Proof. We prove both assertions by induction on n. If $n = 1$, then $R = \mathbb{K}[x_1]$, so when $r = 0$ and $i = 1$, we have $I(K_1)^{\{0\}} = \mathfrak{m} = (x_1)$. Also, we have $b_{1,1}\big(I(K_1)^{\{0\}}\big) = 1$ and $b_{i,j}\big(I(K_1)^{\{0\}}\big) = 0$ for $j \neq i$, which agrees with the expression on the right side of the asserted equality. When $r \geq 1$ or $i \geq 2$, we have $I(K_1)^{\{r\}} = 0$. Since $i + r \geq 2$, we also have $\binom{1}{i+r} = 0$.

Now suppose (by induction) that the formulas hold for $n-1$. If $i \geq 2$, then by Theorem 4.4

$$b_{i,j}(I(K_n)^{\{r\}}) = b_{i,j}(I(K_{n-1})^{\{r\}}) + b_{i-1,j}(I(K_{n-1})^{\{r-1\}}) + b_{i-1,j-1}(I(K_{n-1})^{\{r\}}).$$

If $j \neq i+r$, all terms on the right vanish by the induction hypothesis. If $j = i+r$, the induction hypothesis, in conjunction with the well-known combinatorial identity

$$\binom{m+1}{k+1} = \binom{m}{k+1} + \binom{m}{k}$$

implies

$$
\begin{aligned}
&b_{i,i+r}(I(K_n)^{\{r\}})\\
&\quad=\binom{i+r-1}{r}\binom{n-1}{i+r}+\binom{i+r-2}{r-1}\binom{n-1}{i+r-1}+\binom{i+r-2}{r}\binom{n-1}{i+r-1}\\
&\quad=\binom{i+r-1}{r}\binom{n-1}{i+r}+\left[\binom{i+r-2}{r-1}+\binom{i+r-2}{r}\right]\binom{n-1}{i+r-1}\\
&\quad=\binom{i+r-1}{r}\binom{n-1}{i+r}+\binom{i+r-1}{r}\binom{n-1}{i+r-1}\\
&\quad=\binom{i+r-1}{r}\left[\binom{n-1}{i+r}+\binom{n-1}{i+r-1}\right]\\
&\quad=\binom{i+r-1}{r}\binom{n}{i+r}.
\end{aligned}
$$

Finally, in the case $i=1$, we have

$$b_{1,j}(I(K_n)^{\{r\}})=b_{1,j}(I(K_{n-1})^{\{r\}})+b_{1,j-1}(I(K_{n-1})^{\{r-1\}}).$$

If $j\neq 1+r$, then both terms on the right vanish by induction. If $j=1+r$, the induction hypothesis implies

$$
\begin{aligned}
b_{1,1+r}(I(K_n)^{\{r\}})&=b_{1,1+r}(I(K_{n-1})^{\{r\}})+b_{1,r}(I(K_{n-1})^{\{r-1\}})\\
&=\binom{n-1}{r+1}+\binom{n-1}{r}=\binom{n}{r+1}.
\end{aligned}
$$

This completes the inductive step. □

Complete multipartite graphs. If $m\geq 2$ and $n_1,\dots,n_m$ are positive integers, then the complete multipartite graph $K_{n_1,\dots,n_m}$ may be defined as the m-fold join $\overline{K_{n_1}}\vee\cdots\vee\overline{K_{n_m}}$. It is easily seen that $\chi(K_{n_1,\dots,n_m})=m$. Jacques has computed the Betti numbers of the edge ideal of a complete bipartite graph; since its higher secant powers all vanish, there is nothing more to be done in this case.

Theorem 5.2 [Jacques 2004, Theorem 5.2.4].

$$b_{i,j}(I(K_{n_1,n_2}))=\begin{cases}\sum_{k,\ell\geq 1:k+\ell=i+1}\binom{n_1}{k}\binom{n_2}{\ell} & \text{if } j=i+1,\\ 0 & \text{if } j\neq i+1.\end{cases}$$

If $m\geq 3$, we may realize $K_{n_1,\dots,n_m}$ as $K_{n_1,\dots,n_{m-1}}\vee\overline{K_{n_m}}$ and use Theorem 4.4 to perform a recursive computation, ultimately expressing everything in terms of the quantities appearing in Theorem 5.2. Unfortunately, there does not seem to be a nice formula in closed form. Nevertheless, it is quite easy to establish the following:

Proposition 5.3. *Let $m\geq 2$. If $j\neq i+r$, then $b_{i,j}(I(K_{n_1,\dots,n_m})^{\{r\}})=0$.*

Proof. We proceed by induction on m. The base case ($m = 2$) is Theorem 5.2. Suppose now that the result is known for all positive values $k \le m-1$. If $i \ne 2$, then using Theorem 4.4, we have

$$b_{i,j}(I(K_{n_1,\dots,n_{m-1},1})^{\{r\}}) = b_{i,j}(I(K_{n_1,\dots,n_{m-1}})^{\{r\}}) + b_{i,j-1}(I(K_{n_1,\dots,n_{m-1}})^{\{r-1\}}) \\ + b_{i-1,j-1}(I(K_{n_1,\dots,n_{m-1}})^{\{r\}}).$$

If $j \ne i+r$, then all three terms on the right vanish by induction, and the result holds when $n_m = 1$. Now suppose the result holds when $n_m = k \ge 1$. Then

$$b_{i,j}(I(K_{n_1,\dots,n_{m-1},k+1})^{\{r\}}) = b_{i,j}(I(K_{n_1,\dots,n_{m-1},k})^{\{r\}}) + b_{i,j-1}(I(K_{n_1,\dots,n_{m-1}})^{\{r-1\}}) \\ + b_{i-1,j-1}(I(K_{n_1,\dots,n_{m-1},k})^{\{r\}}).$$

Again, all terms on the right vanish showing that the result holds for $n_m = k+1$. The argument for $i = 1$ is similar. $\square$

We conclude this discussion by giving a clean computation of the simplest nontrivial example in this family — the Betti numbers of the second secant power of the edge ideal of a complete tripartite graph — using a different type of edge-splitting argument. In preparation for the calculation, we introduce a counting function. For $i \ge 1$, $m \ge 2$ and $t \le m$, define

$$P(i,t;n_1,\dots,n_m) = \sum_{\substack{1\le j_1<\cdots<j_t\le m \\ \alpha_1+\cdots+\alpha_t=i+1,\ \alpha_k>0}} \binom{n_{j_1}}{\alpha_1}\cdots\binom{n_{j_t}}{\alpha_t}.$$

If we consider m bins with respective capacities $n_1,\dots,n_m$, the function defined above counts the number of ways of distributing $i+1$ balls among exactly t of these bins.

The Betti numbers of the edge ideal of the complete multipartite graph were also computed by Jacques:

Theorem 5.4 [Jacques 2004, Theorem 5.3.8]. *Suppose $i, m \ge 1$. Then*

$$b_{i,i+1}(I(B_{n_1,\dots,n_m})) = \sum_{t=2}^{m}(t-1)P(i,t;n_1,\dots,n_m).$$

We now have the tools necessary for our calculation.

Proposition 5.5. *Suppose $i \ge 1$. Then*

$$b_{i,i+2}\big(I(B_{n_1,n_2,n_3})^{(2)}\big) = P(i+1,2;n_1,n_2,n_3) + 2P(i+1,3;n_1,n_2.n_3) \\ - P(i+1,2;n_1,n_3) - P(i+1,2;n_2,n_1+n_3).$$

Proof. For convenience, let $I = I(B_{n_1,n_2,n_3}) \subseteq \mathbb{K}[x_1,\dots,x_{n_1},y_1,\dots,y_{n_2},z_1,\dots,z_{n_3}]$, J be the ideal generated by the various products x_iz_k, where $1 \le i \le n_1$ and $1 \le k \le n_3$, and K be the ideal generated by the products x_iy_j and y_jz_k, where $1 \le i \le n_1$,

$1 \le j \le n_2$, and $1 \le k \le n_3$. By Proposition 3.2, $J \cap K$ is generated by the products $x_i y_j z_k$, where i, j, and k are as above. By Theorem 3.1, we see that in fact $I^{\{2\}} = J \cap K$. Furthermore, the map $x_i y_j z_k \mapsto (x_i z_k, x_i y_j)$ is a splitting function, and thus witnesses that $I = J + K$ is a splitting.

By Theorem 3.4, we have

$$b_{i,j}(I^{\{2\}}) = b_{i,j}(J \cap K) = b_{i+1,j}(I) - b_{i+1,j}(J) - b_{i+1,j}(K).$$

Now $I = I(B_{n_1,n_2,n_3})$, $J = I(B_{n_1,n_3})$, and $K = I(B_{n_2,n_1+n_3})$, so by Theorem 5.4, we have

$$\begin{aligned} b_{i,i+2}(I^{\{2\}}) = P(i+1, 2; n_1, n_2, n_3) + 2P(i+1, 3; n_1, n_2.n_3) \\ - P(i+1, 2; n_1, n_3) - P(i+1, 2; n_2, n_1+n_3). \quad \square \end{aligned}$$

The key insight here was to identify the secant ideal as the intersection of two ideals which (along with their sum) are better understood, and to apply the ideal splitting formula in reverse. Unfortunately, this technique does not seem to extend to a more general setting.

Acknowledgements

Part of this work was done at the Summer Undergraduate Mathematical Sciences Research Institute held at Miami University during June and July 2013; it was later augmented and generalized by the first author. The authors also used the software Macaulay2 to aid in the computation of examples. The authors thank the National Security Agency, the National Science Foundation, and Miami University for support during this time. They also thank Hamid Rahmati and Jessica Sidman for helpful discussions and correspondence.

References

[Cranfill 2009] R. Cranfill, "Edge ideals of chordal graphs and generic initial ideals", Mount Holyoke College, South Hadley, MA, 2009, http://www.mtholyoke.edu/~jsidman/reu09PapersandSlides/rachelPaper.pdf.

[Eisenbud 1995] D. Eisenbud, *Commutative algebra: With a view toward algebraic geometry*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR 1322960 Zbl 0819.13001

[Eliahou and Kervaire 1990] S. Eliahou and M. Kervaire, "Minimal resolutions of some monomial ideals", *J. Algebra* **129**:1 (1990), 1–25. MR 1037391 Zbl 0701.13006

[Emtander 2009] E. Emtander, "Betti numbers of hypergraphs", *Comm. Algebra* **37**:5 (2009), 1545–1571. MR 2526320 Zbl 1191.13015

[Fatabbi 2001] G. Fatabbi, "On the resolution of ideals of fat points", *J. Algebra* **242**:1 (2001), 92–108. MR 1844699 Zbl 0984.14016

[Francisco et al. 2009] C. A. Francisco, H. T. Hà, and A. Van Tuyl, "Splittings of monomial ideals", *Proc. Amer. Math. Soc.* **137**:10 (2009), 3271–3282. MR 2515396 Zbl 1180.13018

[Hà and Van Tuyl 2007] H. T. Hà and A. Van Tuyl, "Splittable ideals and the resolutions of monomial ideals", *J. Algebra* **309**:1 (2007), 405–425. MR 2301246 Zbl 1151.13017

[Hà and Van Tuyl 2008] H. T. Hà and A. Van Tuyl, "Monomial ideals, edge ideals of hypergraphs, and their graded Betti numbers", *J. Algebraic Combin.* **27**:2 (2008), 215–245. MR 2375493 Zbl 1147.05051

[Hochster 1977] M. Hochster, "Cohen–Macaulay rings, combinatorics, and simplicial complexes", pp. 171–223 in *Ring theory II* (Norman, OK, 1975), edited by B. R. McDonald and R. A. Morris, Lecture Notes in Pure and Applied Mathematics **26**, Dekker, New York, 1977. MR 0441987 Zbl 0351.13009

[Jacques 2004] S. Jacques, *Betti numbers of graph ideals*, thesis, University of Sheffield, 2004. arXiv math/0410107

[Rosen 2009] Z. Rosen, "Edge and secant ideals of shared-vertex graphs", Mount Holyoke College, South Hadley, MA, 2009, http://www.mtholyoke.edu/~jsidman/reu09PapersandSlides/zviPaper.pdf.

[Sidman and Vermeire 2009] J. Sidman and P. Vermeire, "Syzygies of the secant variety of a curve", *Algebra Number Theory* **3**:4 (2009), 445–465. MR 2525559 Zbl 1169.13304

[Sidman and Vermeire 2011] J. Sidman and P. Vermeire, "Equations defining secant varieties: geometry and computation", pp. 155–174 in *Combinatorial aspects of commutative algebra and algebraic geometry*, edited by G. Fløystad et al., Abel Symposia **6**, Springer, Berlin, 2011. MR 2810430 Zbl 1251.14043

[Simis and Ulrich 2000] A. Simis and B. Ulrich, "On the ideal of an embedded join", *J. Algebra* **226**:1 (2000), 1–14. MR 1749874 Zbl 1034.14026

[Sturmfels and Sullivant 2006] B. Sturmfels and S. Sullivant, "Combinatorial secant varieties", *Pure Appl. Math. Q.* **2**:3, part 1 (2006), 867–891. MR 2252121 Zbl 1107.14045

[Villarreal 1995] R. H. Villarreal, "Rees algebras of edge ideals", *Comm. Algebra* **23**:9 (1995), 3513–3524. MR 1335312 Zbl 0836.13014

[West 1996] D. B. West, *Introduction to graph theory*, Prentice Hall, Upper Saddle River, NJ, 1996. MR 1367739 Zbl 0845.05001

Received: 2014-12-31 Revised: 2015-07-23 Accepted: 2015-10-27

akhtarr@miamioh.edu — *Department of Mathematics, Miami University, Oxford, OH 45056, United States*

brittanynoel@knights.ucf.edu — *Department of Mathematics, University of Central Florida, Orlando, FL 32816, United States*

haleydohrmann@gmail.com — *Department of Mathematics, University of Iowa, Iowa City, IA 52242, United States*

hoganshl@miamioh.edu — *Department of Mathematics, University of Utah, Salt Lake City, UT 84112, United States*

sobieska@math.tamu.edu — *Department of Mathematics, Texas A&M University, College Station, TX 77843, United States*

zwoods@math.uga.edu — *Department of Mathematics, University of Georgia, Athens, GA 30602, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.751

Convergence of sequences of polygons

Eric Hintikka and Xingping Sun

(Communicated by Zuhair Nashed)

In 1878, Darboux studied the problem of midpoint iteration of polygons. Simply put, he constructed a sequence of polygons $\Pi^{(0)}, \Pi^{(1)}, \Pi^{(2)}, \dots$ in which the vertices of a descendant polygon $\Pi^{(k)}$ are the midpoints of its parent polygon $\Pi^{(k-1)}$ and are connected by edges in the same order as those of $\Pi^{(k-1)}$. He showed that such a sequence of polygons converges to their common centroid. In proving this result, Darboux utilized the powerful mathematical tool we know today as the finite Fourier transform. For a long time period, however, neither Darboux's result nor his method was widely known. The same problem was proposed in 1932 by Rosenman as Monthly Problem # 3547 and had been studied by several authors, including I. J. Schoenberg (1950), who also employed the finite Fourier transform technique. In this paper, we study generalizations of this problem. Our scheme for the construction of a polygon sequence not only gives freedom in selecting the vertices of a descendant polygon but also allows the polygon generating procedure itself to vary from one step to another. We show under some mild restrictions that a sequence of polygons thus constructed converges to a single point. Our main mathematical tools are ergodicity coefficients and the Perron–Frobenius theory on nonnegative matrices.

1. Introduction

Jean Gaston Darboux [1878] proposed and solved the following problem. Let $\Pi^{(0)}$ be a closed polygon in the plane with vertices

$$v_0^{(0)}, v_1^{(0)}, \dots, v_{n-1}^{(0)}.$$

Denote by

$$v_0^{(1)}, v_1^{(1)}, \dots, v_{n-1}^{(1)},$$

respectively, the midpoints of the edges $v_0^{(0)}v_1^{(0)},\ v_1^{(0)}v_2^{(0)}, \dots, v_{n-1}^{(0)}v_0^{(0)}$. Connecting $v_0^{(1)}, v_1^{(1)}, \dots, v_{n-1}^{(1)}$ in the same order as above, we derive a new polygon, denoted by $\Pi^{(1)}$. Apply the same procedure to derive polygon $\Pi^{(2)}$. After k constructions, we obtain polygon $\Pi^{(k)}$. Show that $\Pi^{(k)}$ converges, as $k \to \infty$, to the centroid of the

MSC2010: primary 15A60, 65F35; secondary 15A12, 15A30.
Keywords: polygons, finite Fourier transform, stochastic matrices, ergodicity coefficients.

original points $v_0^{(0)}, v_1^{(0)}, \dots, v_{n-1}^{(0)}$. We will refer to this problem as "midpoint iteration of polygons". For any given sequence of polygons $\Pi^{(0)}$, $\Pi^{(1)}$, $\Pi^{(2)}, \dots,$ we will call $\Pi^{(k)}$ the descendant polygon of $\Pi^{(k-1)}$ and $\Pi^{(k-1)}$ the parent polygon of $\Pi^{(k)}$.

In proving his result, Darboux used the powerful mathematical tool we know today as the finite Fourier transform. This allowed him to establish an exponential rate at which the polygon sequence converges to their common centroid. For a long period of time, however, neither Darboux's result nor his method was widely known. More than half a century later, the same problem, which has since been known as Monthly Problem # 3547, was proposed by Rosenman, and a solution of the problem was given by R. Huston in [Rosenman and Huston 1933].

Unaware of what Darboux had already done, Schoenberg [1950] completely retooled the finite Fourier transform technique to tackle the problem of midpoint iteration of polygons. Schoenberg also generalized the problem by allowing vertices of a descendant polygon to come from convex hulls of consecutive vertices of its parent polygon. Later, Schoenberg [1982] revisited this interesting topic. Terras [1999] summarized Schoenberg's work as an example of applications of the finite Fourier transform. One can approach the problem of midpoint iteration of polygons from other mathematical perspectives. For example, Ding et al. [2003] and Ouyang [2013] considered this problem as a special case of Markov chains, and Treatman and Wickham [2000] studied a logarithmic dual problem in which all the vertices of the polygons are on the unit circle and the convergence is to a regular polygon.

In this paper, we study several generalizations of this problem. In Section 3, we consider cases in which the vertices of a descendant polygon are not necessarily midpoints of the edges of its parent polygon but can be chosen more freely from the edges of its parent polygon. In Section 4, we further generalize the work done in Section 3 by allowing the polygon generating procedure to vary from one step to another. In Section 5, we again elevate the level of freedom in selecting the vertices of a descendant polygon by allowing them to come from convex hulls of some subsets of the vertices of its parent polygon. Technically, Section 4 deals with a special case of what is studied in Section 5. In our opinion, however, the importance of the special case deserves some special attention, as does the mathematical argument employed therein. Furthermore, our results in Section 4 are more quantitative, and their geometric implications more illustrative. In addition, the flow of representation reflects the progressive nature of our research process. Section 2 is devoted to the introduction of frequently used notations and definitions.

To conclude the introduction of this paper, we share with readers a few highlights of this research experience. In the midpoint polygon iteration problem, if we view the collection of vertices of a polygon as a vector $z := (z_0, z_1, \dots, z_{n-1})^T$ in $\mathbb{C}^n$, then the collection of vertices of the first descendant polygon is $A\boldsymbol{z}$, where

$$A := \operatorname{circ}\left(\tfrac{1}{2}, \tfrac{1}{2}, 0, \dots, 0\right),$$

in which $\operatorname{circ}\left(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0\right)$ denotes the $n \times n$ circulant matrix[1] whose first row is $\left(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0\right)$. Likewise, the k-th polygon has vertices $A^k\boldsymbol{z}$. Scrutinizing Darboux and Schoenberg's proofs, we found that they had implicitly established the stronger result that $\|A^k - L\|_2$ converges to zero exponentially, where $\|\cdot\|_2$ indicates the spectral radius norm for square matrices, and L is the rank-one matrix whose entries are all $1/n$. It follows that the sequence of the polygons converges to their common centroid. We briefly entertained several possible ways to generalize this problem before we chose to focus on investigating the asymptotic behavior of products of (square) row stochastic matrices and the geometric implications for the corresponding sequence of polygons. Witnessing that the finite Fourier transform works wonderfully with circulant matrices, we tried bounding an arbitrary stochastic matrix by a sum of circulant stochastic matrices. While we have had some success with this strategy in estimating the smallest eigenvalue of a nonsingular stochastic matrix, we have yet to retool the method in a suitable way for the problem in this paper. Our basic tools in this paper are the Perron–Frobenius theorem [Horn and Johnson 1990] on nonnegative matrices and ergodicity coefficients [Ipsen and Selee 2011].

2. Notations and definitions

We will use boldface letters, such as $\boldsymbol{v}$, to denote vectors in $\mathbb{C}^n$. The i-th component of $\boldsymbol{v}$ is denoted by v_i. When the full form of the vector $\boldsymbol{v}$ is needed in some context, we will write $\boldsymbol{v} = (v_0, v_1, \dots, v_{n-1})^T$.

Let n complex numbers (not necessarily distinct) be given. We may connect them in any given order to form a (possibly degenerate) n-gon in the complex plane. In this way, an n-gon can be identified with a vector in $\mathbb{C}^n$, and vice versa. Label the n complex numbers according to the order in which they are connected by edges: $v_0, v_1, \dots, v_{n-1}, v_n, \dots$. That is, two components are adjacent if and only if the corresponding vertices are connected by an edge. To facilitate mathematical exposition, we have here adopted arithmetic modulo n. For example, v_0 and v_n are the same vertex. We will use the same modular arithmetic for row and column indices of matrix entries, announcing as we do so.

If A is a matrix, we denote by $(A)_{ij}$ the entry of A located at the i-th row and the j-th column. If A is a square matrix, then the *spectral radius* of A is denoted by $\rho(A)$, and we define

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

[1] The definition of circulant matrices will be given in Section 2.

Definition 1. A matrix A is *positive*, denoted by $A > 0$, if $(A)_{ij} > 0$ for all i, j. Similarly, A is *nonnegative*, or $A \geq 0$, if $(A)_{ij} \geq 0$ for all i, j. If, for some $\alpha \in \mathbb{R}$, we have $(A)_{ij} > \alpha$ (respectively, $(A)_{ij} \geq \alpha$) for all i, j, then we will write $A > \alpha$ (respectively, $A \geq \alpha$).

Definition 2. A *stochastic* (or *row-stochastic*) matrix is a real-valued, nonnegative, square matrix whose row sums are all 1.

Definition 3. An $n \times n$ matrix A is *circulant* if for some complex numbers a_i, we have

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-3} & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_2 & a_3 & \cdots & a_0 & a_1 \\ a_1 & a_2 & \cdots & a_{n-1} & a_0 \end{bmatrix}.$$

In Section 1, we used the notation $\operatorname{circ}(a_0, a_1, \ldots, a_{n-2}, a_{n-1})$ to denote the above circulant matrix. We will continue to do so in appropriate contexts.

Definition 4. We call an $n \times m$ matrix A *k-banded* if

$$(A)_{ij} \neq 0 \iff j \in \{i, i+1, \ldots, i+k-1\} \pmod n.$$

For example, the matrix $\operatorname{circ}(1/2, 1/2, 0, \ldots, 0)$ is a 2-banded matrix.

Definition 5. We say that two $n \times m$ matrices A and B have the same *zero pattern* if $(A)_{ij} = 0 \iff (B)_{ij} = 0$ for all i, j.

Definition 6. We say that an $n \times n$ matrix is *circulant-patterned* if it has the same zero pattern as a circulant matrix.

Definition 7. We say that a sequence of n-gons $(\Pi^{(k)})_{k \geq 0}$ *converges to a point $q \in \mathbb{C}$* if, for any $1 \leq p \leq \infty$, we have $\lim_{k \to \infty} \|\Pi^{(k)} - \boldsymbol{q}\|_p = 0$, where $\boldsymbol{q} = (q, q, \ldots, q)^T$. Here $\|\cdot\|_p$ denotes the p-norm on $\mathbb{C}^n$, that is,

$$\|\boldsymbol{v}\|_p = \begin{cases} \left(\sum_{i=1}^n |v_i|^p\right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \max\limits_{1 \leq i \leq n} |v_i| & \text{if } p = \infty. \end{cases}$$

In this paper, we primarily work with the 1-norm. To be sure, any two norms on a finite-dimensional normed linear space are topologically equivalent.

3. Polygons derived from a fixed 2-banded matrix

In this section, we suppose that $\Pi^{(0)}$ is an n-gon and that its k-th descendant polygon is given by $\Pi^{(k)} = A^k \Pi^{(0)}$, where A is a fixed 2-banded stochastic matrix. Hence, for some real $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ with $0 < \alpha_i < 1$, we have

$$A = \begin{bmatrix} \alpha_0 & 1-\alpha_0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_1 & 1-\alpha_1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_{n-1} & 0 & \cdots & 0 & 0 & \alpha_{n-1} \end{bmatrix}. \tag{3-1}$$

We are interested in this particular construction because, geometrically speaking, the vertices of $\Pi^{(k+1)}$ are chosen from within the edges of $\Pi^{(k)}$, one vertex per edge. The procedure also allows the choice of any particular vertex in $\Pi^{(k+1)}$ to be independent from the others. We aim to show that $(\Pi^{(k)})_{k\geq 0}$ converges to a predetermined point. In the present section, the theoretic foundation for our argument is Perron's theorem (8.2.11(f) in [Horn and Johnson 1990]), which we state in the following theorem. To be sure, the convergence results in this section follow from the general framework of the Perron–Frobenius theorem. However, the 2-banded structure of our matrices allows us to obtain more nuanced convergence results. In particular, our knowledge of the convergence process is quantitative in the sense that we are able to predetermine the point to which the sequence of polygons converges.

Theorem (Perron). *If A is a positive $n \times n$ matrix, then*

$$[\rho(A)^{-1}A]^m \to L \quad \text{as } m \to \infty,$$

where $L = xy^T$, $Ax = \rho(A)x$, $A^T y = \rho(A)y$, $x > 0$, $y > 0$, *and* $x^T y = 1$.

We now derive some quick results and use these, along with Perron's theorem, to show that the sequence $(\Pi^{(k)})_{k\geq 0}$ in fact converges to a point for any choice of A. Additionally, we give an expression for that limiting point in terms of the entries of A.

Proposition 8. *If A and A_i, where $i \in \{0, 1, \dots, k-1\}$, are $n \times n$ stochastic matrices, then we have*:

(1) *The spectral radius $\rho(A)$ is* 1.

(2) *The product matrix $A_{k-1}A_{k-2}\cdots A_0$ is stochastic.*

Proof. These are known results. Part (1) follows from Lemma 8.1.21 in [Horn and Johnson 1990]. We give a short yet entertaining proof to part (2) using the simple fact that an $n \times n$ matrix A is stochastic if and only if $Ae = e$, where $e \in \mathbb{R}^n$ is the vector with all components 1. We simply write

$$A_{k-1}A_{k-2}\cdots A_0 e = A_{k-1}A_{k-2}\cdots A_1 e = \cdots = e. \qquad \square$$

Proposition 9. *Suppose that A is a 2-banded stochastic matrix as given in (3-1). Then $A^{n-1} > 0$.*

This result is stated in [Ouyang 2013] without a proof. We give a complete proof here, as variations of it will become quite useful in the latter part of the paper.

Proof. Throughout the proof, we use arithmetic modulo n to track the changes in row and column indices as a result of matrix multiplications. For $n = 2$, the result is obvious. Suppose that for some $N \in \mathbb{N}$, we have $(A^N)_{ij} > 0$ for all i and j such that $j \in \{i, i+1, \dots, i+N\} \pmod n$. Then for any such i and j, we have

$$(A^{N+1})_{ij} = \sum_{k=0}^{n-1} (A^N)_{ik}(A)_{kj} \geq (A^N)_{ij}(A)_{jj} > 0.$$

Furthermore, we have

$$\begin{aligned}(A^{N+1})_{i,i+N+1} &= \sum_{k=0}^{n-1} (A)_{ik}(A^N)_{k,i+N+1} \\ &= \alpha_i (A^N)_{i,i+N+1} + (1-\alpha_i)(A^N)_{i+1,i+N+1} \\ &\geq (1-\alpha_i)(A^N)_{i+1,i+N+1},\end{aligned}$$

which is positive by the induction hypothesis. It follows that the matrix A^{N+1} has positive entries at (i, j) whenever $j \in \{i, i+1, \dots, i+N+1\} \pmod n$. Hence A^{n-1} has positive entries everywhere. □

Proposition 10. *Let A be a matrix as given in (3-1). Then we have*

$$\lim_{k \to \infty} A^k = L,$$

where L is the rank-one matrix given by (3-2) in the proof below.

Proof. Let $B = A^{n-1}$. Since B is the product of $(n-1)$ stochastic matrices, it is itself stochastic by Proposition 8. Furthermore, we have that $\rho(B) = 1$. Let

$$y = ((1-\alpha_0)^{-1}, (1-\alpha_1)^{-1}, \dots, (1-\alpha_{n-1})^{-1})^T.$$

One can verify that $A^T y = y$. Thus, $B^T y = (A^{n-1})^T y = (A^T)^{n-1} y = y = \rho(B) y$. Let $x = \alpha_A (1, \dots, 1)^T$, where α_A is the scalar given by $\alpha_A = \bigl(\sum_{i=0}^{n-1} (1-\alpha_i)^{-1}\bigr)^{-1}$. Then we have that $Bx = \rho(B)x$ and that $x^T y = 1$. Let $L = xy^T$. Since $x > 0$ and $y > 0$, the rank-one matrix L has identical rows. Specifically,

$$L = \begin{bmatrix} \Bigl((1-\alpha_0)\sum_{i=0}^{n-1}\frac{1}{1-\alpha_i}\Bigr)^{-1} & \cdots & \Bigl((1-\alpha_{n-1})\sum_{i=0}^{n-1}\frac{1}{1-\alpha_i}\Bigr)^{-1} \\ \vdots & & \vdots \\ \Bigl((1-\alpha_0)\sum_{i=0}^{n-1}\frac{1}{1-\alpha_i}\Bigr)^{-1} & \cdots & \Bigl((1-\alpha_{n-1})\sum_{i=0}^{n-1}\frac{1}{1-\alpha_i}\Bigr)^{-1} \end{bmatrix}. \tag{3-2}$$

By Proposition 9, we have $B > 0$. Applying Perron's theorem, we conclude that

$$\lim_{k \to \infty} B^k = L.$$

The rest of the proof is devoted to showing that $\lim_{k\to\infty} A^k = L$. Since L has identical rows and A^i is a stochastic matrix for any $i \in \{0, 1, \dots, n-2\}$, we have, for all l, m, that

$$(A^i L)_{lm} = \sum_{k=0}^{n-1} (A^i)_{lk}(L)_{km} = \sum_{k=0}^{n-1} (A^i)_{lk}(L)_{lm} = (L)_{lm} \sum_{k=0}^{n-1} (A^i)_{lk} = (L)_{lm},$$

that is, $L = A^i L$. Hence we have

$$L = A^i \lim_{k\to\infty} B^k = \lim_{k\to\infty} A^i A^{k(n-1)} = \lim_{k\to\infty} A^{k(n-1)+i} \quad \text{for } 0 \le i \le n-2.$$

For a given $\epsilon > 0$, let N_i be such that $\|A^{m(n-1)+i} - L\| < \epsilon$ for all $m \ge N_i$. Let $N = \max\{N_i : i \in \{0, 1, \dots, n-2\}\}$. Choose $j \ge N(n-1)+(n-2)$. By the division theorem, $j = m(n-1)+i$ for some integer m and some fixed $i \in \{0, 1, \dots, n-2\}$. So,

$$j = m(n-1) + i \ge N(n-1) + (n-2) \ge N(n-1) + i \ge N_i(n-1) + i.$$

Hence $m \ge N_i$. Thus we have $\|A^j - L\| < \epsilon$. This inequality holds true for all $j \ge N(n-1) + (n-2)$, which proves that $\lim_{k\to\infty} A^k = L$. □

As the main theorem of this section, we restate the result of Proposition 10 in terms of convergence of a sequence of polygons.

Theorem 11. *Let $(\Pi^{(k)})_{k\ge 0}$ be a polygon sequence constructed by $\Pi^{(k)} = A^k \Pi^{(0)}$, where A is given as in (3-1). Then we have*

$$\lim_{k\to\infty} \Pi^{(k)} = (q, \dots, q)^T,$$

where

$$q = \sum_{j=0}^{n-1} \left((1-\alpha_j) \sum_{i=0}^{n-1} \frac{1}{1-\alpha_i} \right)^{-1} \Pi_j^{(0)}.$$

We remind readers that a matrix A given as in (3-1) is circulant if and only if $\alpha_i = \alpha_j$ for all i, j. When this holds true, we have

$$(q, \dots, q)^T = \tfrac{1}{n}(\Pi_0^{(0)} + \Pi_1^{(0)} + \dots + \Pi_{n-1}^{(0)}),$$

which is the centroid of the vertices of $\Pi^{(0)}$. The special case that $\alpha_i = \frac{1}{2}$ for all i corresponds to the problem of midpoint iteration of polygons.

4. Polygons derived from a sequence of 2-banded matrices

Let $\delta \in \left(0, \frac{1}{2}\right)$. For each $k \in \mathbb{N}$, we arbitrarily choose n numbers $\alpha_0^{(k)}, \alpha_1^{(k)}, \dots, \alpha_{n-1}^{(k)}$ from the open interval $(\delta, 1-\delta)$ and form the matrix

$$A_k = \begin{bmatrix} \alpha_0^{(k)} & 1-\alpha_0^{(k)} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_1^{(k)} & 1-\alpha_1^{(k)} & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_{n-1}^{(k)} & 0 & \cdots & 0 & 0 & \alpha_{n-1}^{(k)} \end{bmatrix}. \tag{4-1}$$

Let $\mathcal{A}_k = A_k A_{k-1} \cdots A_1$. Let $\Pi^{(0)}$ be an n-gon, and define $\Pi^{(k)} = \mathcal{A}_k \Pi^{(0)}$. We will show that the sequence of polygons $(\Pi^{(k)})_{k \geq 1}$ converges to a point. Under these circumstances, Perron's theorem is no longer applicable. Our argument relies on some key properties of ergodicity coefficients thoroughly studied in a recent article by Ipsen and Selee [2011].

Definition 12. The 1-norm ergodicity coefficient $\tau_1(S)$ for an $n \times n$ stochastic matrix S is given by

$$\tau_1(S) = \max_{\substack{\|z\|_1 = 1 \\ z^T e = 0}} \|S^T z\|_1,$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^n$ and the maximum ranges over $z \in \mathbb{R}^n$. If $n = 1$, we say that $\tau_1(S) = 0$.

Proposition 13. *If S, S_1, and S_2 are stochastic matrices, then*:

(1) $0 \leq \tau_1(S) \leq 1$. *Furthermore*, $\tau_1(S) = 0 \Longleftrightarrow S$ *is a rank-one matrix.*

(2) $|\lambda| \leq \tau_1(S)$ *for all eigenvalues* $\lambda < 1$ *of* S.

(3) $\tau_1(S) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |(S)_{ik} - (S)_{jk}|$.

(4) $\tau_1(S_1 S_2) \leq \tau_1(S_1)\tau_1(S_2)$.

Proof. Part (1) is from Theorem 3.4 in [Ipsen and Selee 2011], while parts (2) and (4) are the results of Theorem 3.6, and part (3) is the result of Theorem 3.7, of the same work. □

Ergodicity coefficients can be defined and studied for all p-norms and even more general metrics under broad matrix analysis settings. For our purpose, however, the results in Proposition 13 suffice.

To proceed, we need the following generalization of Proposition 9.

Proposition 14. *For* $1 \leq k \leq n-1$, *let* A_k *be as defined in* (4-1). *Then*

$$\mathcal{A}_{n-1} > \delta^{n-1}.$$

Proof. The proof can be considered as a quantification of that of Proposition 9. Arithmetic modulo n will be used to track the changes in row and column indices stemming from matrix multiplications. Suppose that for some $N \in \{1, 2, \ldots, n-2\}$ we have $(\mathcal{A}_N)_{ij} > \delta^N$ for all $i \in \{1, 2, \ldots, n\}$ and $j \in \{i, i+1, \ldots, i+N\}$. Then, for all such i and j,

$$\begin{aligned}(\mathcal{A}_{N+1})_{ij} &= \sum_{k=1}^{n} (A_{N+1})_{ik} (\mathcal{A}_N)_{kj} \\ &\geq (A_{N+1})_{ii} (\mathcal{A}_N)_{ij} > \delta \cdot \delta^N = \delta^{N+1},\end{aligned}$$

that is, $(\mathcal{A}_{N+1})_{ij} > \delta^{N+1}$. Also,

$$\begin{aligned}(\mathcal{A}_{N+1})_{i,(i+N+1)} &= \alpha_i^{(N+1)} \cdot (\mathcal{A}_N)_{i,(i+N+1)} + (1-\alpha_i^{(N+1)}) \cdot (\mathcal{A}_N)_{(i+1),(i+N+1)} \\ &\geq (1-\alpha_i^{(N+1)}) \cdot (\mathcal{A}_N)_{(i+1),(i+N+1)} > \delta \cdot \delta^N = \delta^{N+1}.\end{aligned}$$

Thus $(\mathcal{A}_{N+1})_{ij} > \delta^{N+1}$ for all $i \in \{1, 2, \dots, n\}$ and $j \in \{i, i+1, \dots, i+N+1\}$.

Since in the case $N = 1$ it is clearly true that $(\mathcal{A}_N)_{ij} > \delta^N$ for all $i \in \{1, 2, \dots, n\}$ and $j \in \{i, (i \mod n) + 1\}$, it follows from the principle of mathematical induction that $(\mathcal{A}_{n-1})_{ij} > \delta^{n-1}$ for all $i, j \in \{1, 2, \dots, n\}$. □

Proposition 15. *If S is a positive $n \times n$ stochastic matrix and*

$$\epsilon := \min_{i,j}(S)_{ij},$$

then

$$\tau_1(S) \leq 1 - n\epsilon.$$

Proof. We first point out that under the conditions specified in Proposition 15, we have $n\epsilon \leq 1$. Therefore, the number on the right-hand side of the above inequality is nonnegative. Let S_0 be the $n \times n$ matrix defined by $(S_0)_{ij} = (S)_{ij} - \epsilon$ for all i, j. Then S_0 is nonnegative, and the row sums of S_0 are all $1 - n\epsilon$. More pertinently, we have $\tau_1(S) = \tau_1(S_0)$. To calculate $\tau_1(S_0)$, we write, for all $i, j \in \{1, 2, \dots, n\}$, that

$$\sum_{k=1}^{n} |(S_0)_{ik} - (S_0)_{jk}| \leq \sum_{k=1}^{n} |(S_0)_{ik}| + \sum_{k=1}^{n} |(S_0)_{jk}| \leq 2(1 - n\epsilon).$$

The desired result then follows from part (3) of Proposition 15. □

We state our main result of this section in the following theorem.

Theorem 16. *Let A_ℓ $(0 \leq \ell < \infty)$ be a sequence of matrices as given in (4-1), and let $\mathcal{A}_k = A_0 A_1 \cdots A_k$. Then we have*

$$\lim_{k \to \infty} \mathcal{A}_k = L,$$

where L is a rank-one stochastic matrix with identical rows. Hence if $\Pi^{(k)}$ is the corresponding sequence of polygons, we have

$$\lim_{k \to \infty} \Pi^{(k)} = L\Pi^{(0)},$$

and thus $(\Pi^{(k)})_{k \geq 0}$ converges to a point.

Proof. Let $(A_\ell)_{\ell \geq 0}$ be matrices as given in (4-1). For each $k \geq 1$, define

$$\mathcal{B}_k = A_{kn-1} A_{kn-2} \cdots A_{(k-1)n}.$$

Then by Propositions 14 and 15, we have

$$\tau_1(\mathcal{B}_k) \leq 1 - n\delta^{n-1} \quad \text{for } k \geq 1.$$

For a given $k \geq 1$, let $m = \max\{j : jn-1 \leq k\}$. By parts (1) and (4) of Proposition 13, we have

$$\begin{aligned}\tau_1(\mathcal{A}_k) &= \tau_1(A_k A_{k-1} \cdots A_{mn} \mathcal{B}_m \mathcal{B}_{m-1} \cdots \mathcal{B}_1) \\ &\leq \tau_1(\mathcal{B}_m \mathcal{B}_{m-1} \cdots \mathcal{B}_1) \\ &\leq \tau_1(\mathcal{B}_m)\tau_1(\mathcal{B}_{m-1}) \cdots \tau_1(\mathcal{B}_1) \\ &\leq (1 - n\delta^{n-1})^m.\end{aligned}$$

Note that $m \to \infty$ when k does. Thus we have

$$\lim_{k\to\infty} \tau_1(\mathcal{A}_k) \leq \lim_{k\to\infty} (1 - n\delta^{n-1})^m = 0.$$

It follows from part (1) of Proposition 13 that $\mathcal{A}_k$ converges to a rank-one matrix. To show that L has identical rows, we use a Cauchy sequence argument. For any given $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $m \geq N$, we have

$$\frac{1}{2} \max_{i,j} \sum_{l=1}^{n} |(\mathcal{A}_m)_{il} - (\mathcal{A}_m)_{jl}| < \frac{\epsilon}{4}.$$

This implies that

$$\max_{i,j} |(\mathcal{A}_m)_{ij} - (\mathcal{A}_m)_{1j}| < \frac{\epsilon}{2}. \tag{4-2}$$

This allows us to write

$$(\mathcal{A}_m)_{ij} = a_j + \delta_{ij}^{(m)} \quad \text{for } 1 \leq i, j \leq n,$$

in which a_j is fixed for each $1 \leq j \leq n$, and

$$|\delta_{ij}^{(m)}| \leq \frac{\epsilon}{2} \quad \text{for } 1 \leq j \leq n \text{ and } m > N.$$

Upon writing $\mathcal{A}_{m+k} = S_k \mathcal{A}_m$, where S_k is a stochastic matrix, we have

$$\begin{aligned}(\mathcal{A}_{m+k})_{ij} = \sum_{l=1}^{n} (S_k)_{il} (\mathcal{A}_m)_{lj} &= \sum_{l=1}^{n} (S_k)_{il} (a_j + \delta_{lj}^{(m)}) \\ &= a_j \sum_{l=1}^{n} (S_k)_{il} + \sum_{l=1}^{n} (S)_{il} \delta_{lj}^{(m)} \\ &= a_j + \sum_{l=1}^{n} (S_k)_{il} \delta_{lj}^{(m)}.\end{aligned}$$

We also have that

$$-\frac{\epsilon}{2} = -\frac{\epsilon}{2} \sum_{l=1}^{n} (S_k)_{il} < \sum_{l=1}^{n} (S_k)_{il} \delta_{lj}^{(m)} < \frac{\epsilon}{2} \sum_{l=1}^{n} (S_k)_{il} = \frac{\epsilon}{2}.$$

Hence we have $|(\mathcal{A}_{m+k})_{ij} - a_j| < \epsilon/2$, that is,

$$|(\mathcal{A}_{m+k})_{ij} - (\mathcal{A}_m)_{1j}| < \frac{\epsilon}{2}. \tag{4-3}$$

We combine inequalities (4-2) and (4-3) to have

$$\begin{aligned}|(\mathcal{A}_{m+k})_{ij} - (\mathcal{A}_m)_{i'j}| &= |(\mathcal{A}_{m+k})_{ij} - (\mathcal{A}_m)_{1j} + (\mathcal{A}_m)_{1j} - (\mathcal{A}_m)_{i'j}| \\ &\le |(\mathcal{A}_{m+k})_{ij} - (\mathcal{A}_m)_{1j}| + |(\mathcal{A}_m)_{1j} - (\mathcal{A}_m)_{i'j}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,\end{aligned}$$

which is true for all $m > N$, $k \ge 0$, and all $0 \le i, i', j \le n-1$. The above inequality shows that for each fixed i and j, the sequence $(\mathcal{A}_k)_{ij}$ is Cauchy, and that for each fixed $0 \le j \le n-1$, the limits of the sequences $(\mathcal{A}_k)_{ij}$ are the same for all $0 \le i \le n-1$. Thus, for each $0 \le j \le n-1$, there exists a real number q_j such that $\lim_{k\to\infty} (\mathcal{A}_k)_{ij} = q_j$ for all $0 \le i \le n-1$. Hence $\mathcal{A}_k$ converges to the rank-one matrix

$$\begin{bmatrix} q_0 & q_1 & \cdots & q_{n-1} \\ q_0 & q_1 & \cdots & q_{n-1} \\ \vdots & \vdots & & \vdots \\ q_0 & q_1 & \cdots & q_{n-1} \end{bmatrix}. \qquad \square$$

In the above proof, the ij-entries of the rank-one matrix L are given as limits of the sequences $(\mathcal{A}_k)_{ij}$. Since they determine the position where the sequence of the polygons converges, a certain effort should be devoted to finding the limits. Doing so, however, would have gone beyond the scope of this paper.

5. Polygons derived from a sequence of circulant-patterned matrices

In the previous two sections, we were concerned specifically with polygons derived from sequences of 2-banded stochastic matrices. Each descendant polygon thus generated is inscribed in its parent polygon, a fact which may be utilized to control polygons of other types. In this section, we broaden our scope and consider polygons derived from sequences of matrices of a more general class, namely, stochastic circulant-patterned matrices.

Proposition 17. *Suppose that $(A_\ell)_{\ell\ge 0}$ and $(B_\ell)_{\ell\ge 0}$ are two sequences of nonnegative $n \times n$ matrices such that A_ℓ and B_ℓ have the same zero pattern for each ℓ. Then for each $k \in \mathbb{N}$, the two matrices $\mathcal{A}_k = A_{k-1}A_{k-2}\cdots A_0$ and $\mathcal{B}_k = B_{k-1}B_{k-2}\cdots B_0$ share a zero pattern for any k.*

Proof. The proof is by induction. Suppose that $\mathcal{A}_k$ and $\mathcal{B}_k$ have the same zero pattern from some k. For any i,j, we know that $(\mathcal{A}_{k+1})_{ij} = 0$ if and only if, for each l, either $(A_k)_{il} = 0$ or $(\mathcal{A}_k)_{lj} = 0$. But this is the case if and only if $(B_k)_{il} = 0$ or $(\mathcal{B}_k)_{lj} = 0$ for each l, i.e., if and only if $(\mathcal{B}_{k+1})_{ij} = 0$. Since A_0 and B_0 have a common zero pattern, the result follows. $\square$

Proposition 18. *Let $k \in \mathbb{N}$ be given. For each $\ell \in \{0, 1, \ldots, k-1\}$, let A_ℓ be a nonnegative $n \times n$ matrix, and let $\mathcal{A}_{\ell+1} = A_\ell A_{\ell-1} \cdots A_0$. Assume that both sets $\{(A_\ell)_{ij} : (A_\ell)_{ij} > 0,\ 0 \le \ell \le k-1\}$ and $\{(\mathcal{A}_k)_{ij} : (\mathcal{A}_k)_{ij} > 0\}$ are nonempty. Let*

$$\epsilon := \min_{i,j,\ell}\{(A_\ell)_{ij} : (A_\ell)_{ij} > 0,\ 0 \le \ell \le k-1\}.$$

Then the following inequality holds true:

$$\min_{i,j}\{(\mathcal{A}_k)_{ij} : (\mathcal{A}_k)_{ij} > 0\} \ge \epsilon^k.$$

Proof. We again prove by induction. The result is obviously true for $k = 1$. Now suppose that $k > 1$ and that for some $\ell < k$ we have $(\mathcal{A}_\ell)_{lm} \neq 0 \implies (\mathcal{A}_i)_{lm} \ge \epsilon^i$. We write down the lm-entry of the matrix $(\mathcal{A}_{\ell+1})$:

$$(\mathcal{A}_{\ell+1})_{lm} = \sum_{j=1}^{n} (A_{\ell+1})_{lj} (\mathcal{A}_\ell)_{jm}.$$

If $(\mathcal{A}_{\ell+1})_{lm}$ is positive, then there exists a j such that both $(A_{\ell+1})_{lj}$ and $(\mathcal{A}_\ell)_{jm}$ are positive. Since $(A_{\ell+1})_{lj} \ge \epsilon$ and $(\mathcal{A}_\ell)_{jm} \ge \epsilon^\ell$, we have $(\mathcal{A}_{\ell+1})_{lm} \ge \epsilon^{\ell+1}$. That is, $(\mathcal{A}_{\ell+1})_{lm} \neq 0 \implies (\mathcal{A}_{\ell+1})_{lm} \ge \epsilon^{\ell+1}$. The induction process is complete. □

The following result is due to Tollisen and Lengyel [2008].

Proposition 19. *Let A be an $n \times n$ circulant matrix with first row $(c_0, c_1, \ldots, c_{n-1})$. Let $L = \{i : c_i > 0\}$, $u = \min L$, $L' = \{i - u : c_i > 0\}$, and $g = \gcd(L')$. Then*

$$(A^k)_{ij} \approx \begin{cases} \frac{1}{n}\gcd(n, g) & \textit{if } j - i \equiv ku \pmod{\gcd(n, g)}, \\ 0 & \textit{otherwise} \end{cases}$$

as $k \to \infty$.

The rest of this section is devoted to statements and proofs of the main result.

Proposition 20. *Let A be a stochastic circulant matrix such that the sequence A^k converges to a rank-one matrix L. Let $(A_\ell)_{\ell \ge 0}$ be a sequence of stochastic matrices having the same zero pattern as A. Moreover, assume that there exists an $\epsilon > 0$ such that*

$$\min_{i,j,\ell}\{(A_\ell)_{ij} : (A_\ell)_{ij} > 0\} \ge \epsilon.$$

Then the sequence of matrices $A_k A_{k-1} \cdots A_1 A_0$ converges to a rank-one matrix L' with identical rows.

Proof. A result from [Kra and Simanca 2012] asserts that the product of circulant matrices is circulant. Hence A^k is a sequence of stochastic circulant matrices, and so is their limit L. Proposition 19 assures us that each entry of L is either zero or $\gcd(n, g)/n$. Suppose that for some i and j, we have $(L)_{ij} = 0$. Since L is stochastic, we must have $(L)_{ij'} > 0$ for some $j' \neq j$. Using arithmetic modulo n to denote row and column indices, we can identify an ℓ such that $(L)_{i\ell} = 0$ and $(L)_{i,\ell+1} > 0$. Since L is circulant, we have $(L)_{(i+1),(\ell+1)} = (L)_{i\ell} = 0$. Therefore, the i-th and the $(i+1)$-th rows of L must be linearly independent. This contradicts the fact that L is rank-one. Therefore the entries $(L)_{ij}$ are either all zero or all equal to $\gcd(n, g)/n$. That L is stochastic rules out the former. In fact, all entries $(L)_{ij}$ are $1/n$, which implies that $\gcd(n, g) = 1$. It follows that, for k sufficiently large, $A^k > 0$.

Define $\mathcal{B}_\ell = A_{\ell k-1} A_{\ell k-2} \cdots A_{(\ell-1)k}$ for $\ell \geq 1$. Then by Proposition 17, each $\mathcal{B}_\ell$ has the same zero pattern as A^k. That is, $\mathcal{B}_\ell > 0$ for all ℓ. Furthermore, by Proposition 18 we know that $\mathcal{B}_\ell \geq \epsilon^k$. By Proposition 15, we have that

$$\tau_1(\mathcal{B}_\ell) \leq 1 - n\epsilon^k \quad \text{for } \ell \geq 0.$$

It follows that

$$\begin{aligned}\tau_1(A_\ell A_{\ell-1} \cdots A_0) &\leq \tau_1(\mathcal{B}_\ell \mathcal{B}_{\ell-1} \cdots \mathcal{B}_1) \\ &\leq \tau_1(\mathcal{B}_\ell)\tau_1(\mathcal{B}_{\ell-1}) \cdots \tau_1(\mathcal{B}_1) \leq (1 - n\epsilon^k)^\ell,\end{aligned}$$

which implies that the sequence $A_k A_{k-1} \cdots A_1 A_0$ converges to a rank-one matrix L'. Moreover, we can use the same Cauchy sequence argument as in the proof of Theorem 16 to show that the matrix L' has identical rows. □

The following result is worth mentioning.

Proposition 21. *If A is a stochastic circulant matrix, then A^k converges to a rank-one matrix as $k \to \infty$ if and only if* $\gcd(n, g) = 1$.

Proof. On the one hand, as we observed in the previous proof, if A is a stochastic circulant matrix such that A^k converges to a rank-one matrix L as $k \to \infty$, then L must be strictly positive, and hence $\gcd(n, g) = 1$. On the other hand, if A is a stochastic circulant matrix such that $\gcd(n, g) = 1$, then $j - i \equiv ku \pmod{\gcd(n, g)}$. Thus, $\lim_{k\to\infty}(A^k)_{ij} = 1/n$. That is, A^k converges to the rank-one matrix whose entries are all $1/n$ as $k \to \infty$. □

We state the main result of this section in terms of convergent sequences of polygons.

Theorem 22. *Suppose that $(A_\ell)_{\ell\geq 0}$ is a sequence of stochastic, circulant-patterned, $n \times n$ matrices that all have a common zero pattern. Let* $\gcd(n, g) = 1$, *where* $u = \min\{i : a_i > 0\}$ *and* $g = \gcd\{i - u : a_i > 0\}$. *Here, $(a_0, a_1, \ldots, a_{n-1})$ is the first row of A_0. Assume that there exists an $\epsilon > 0$ such that*

$$\min_{i,j,k}\{(A_k)_{ij} : (A_k)_{ij} > 0\} \geq \epsilon.$$

Then the sequence of matrices $A_k A_{k-1} \cdots A_0$ converges to a rank-one matrix L that has identical rows. Hence, if $\Pi^{(k)} = A_k A_{k-1} \cdots A_1 \Pi^{(0)}$, then

$$\lim_{k\to\infty} \Pi^{(k)} = L\Pi^{(0)}.$$

That is, the sequence of polygons $(\Pi^{(k)})_{k\geq 0}$ converges to the point $L\Pi^{(0)}$.

Acknowledgment

The authors thank Alvin Moon, University of California, Davis, for calling their attention to G. Darboux's work, which predated Schoenberg's by more than 70 years.

References

[Darboux 1878] G. Darboux, "Sur un problème de gèométrie élémentaire", *Bull. Sci. Math. Astron.* (2) **2**:1 (1878), 298–304.

[Ding et al. 2003] J. Ding, L. R. Hitt, and X.-M. Zhang, "Markov chains and dynamic geometry of polygons", *Linear Algebra Appl.* **367** (2003), 255–270. MR Zbl

[Horn and Johnson 1990] R. A. Horn and C. R. Johnson, *Matrix analysis*, corrected reprint of 1st ed., Cambridge University Press, 1990. Zbl

[Ipsen and Selee 2011] I. C. F. Ipsen and T. M. Selee, "Ergodicity coefficients defined by vector norms", *SIAM J. Matrix Anal. Appl.* **32**:1 (2011), 153–200. MR Zbl

[Kra and Simanca 2012] I. Kra and S. R. Simanca, "On circulant matrices", *Notices Amer. Math. Soc.* **59**:3 (2012), 368–377. MR Zbl

[Ouyang 2013] C. Ouyang, "A problem concerning the dynamic geometry of polygons", unpublished manuscript, 2013.

[Rosenman and Huston 1933] M. Rosenman and R. E. Huston, "3547", *The American Mathematical Monthly* **40**:3 (1933), 184–185. MR

[Schoenberg 1950] I. J. Schoenberg, "The finite Fourier series and elementary geometry", *Amer. Math. Monthly* **57** (1950), 390–404. MR Zbl

[Schoenberg 1982] I. J. Schoenberg, *Mathematical time exposures*, Mathematical Association of America, Washington, DC, 1982. MR Zbl

[Terras 1999] A. Terras, *Fourier analysis on finite groups and applications*, London Mathematical Society Student Texts **43**, Cambridge University Press, 1999. MR Zbl

[Tollisen and Lengyel 2008] G. P. Tollisen and T. Lengyel, "Intermediate and limiting behavior of powers of some circulant matrices", *Ars Combin.* **88** (2008), 229–255. MR Zbl

[Treatman and Wickham 2000] S. Treatman and C. Wickham, "Constructible approximations of regular polygons", *Amer. Math. Monthly* **107**:10 (2000), 911–922. MR Zbl

Received: 2015-03-04 Accepted: 2015-09-17

ehintikka@rice.edu Department of Mathematics, MS 136, Rice University, P. O. Box 1892, Houston, TX 77005, United States

xsun@missouristate.edu Department of Mathematics, Missouri State University, Springfield, MO 65897, United States

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.765

On the Chermak–Delgado lattices of split metacyclic p-groups

Erin Brush, Jill Dietz, Kendra Johnson-Tesch and Brianne Power

(Communicated by Kenneth S. Berenhaut)

The Chermak–Delgado measure of a subgroup H of a finite group G is defined as $m_G(H) = |H||C_G(H)|$. The subgroups with maximal Chermak–Delgado measure form a poset and corresponding lattice, known as the CD-lattice of G. We describe the symmetric nature of CD-lattices in general, and use information about centrally large subgroups to determine the CD-lattices of split metacyclic p-groups in particular. We also describe a rank-symmetric sublattice of the CD-lattice of split metacyclic p-groups.

1. Introduction

A. Chermak and A. Delgado [1989] developed a "measuring argument" for a finite group G acting on a finite group H to prove, as G. Glauberman [2006] put it, "remarkably beautiful results and powerful applications." Glauberman extended Chermak and Delgado's work to obtain, among other things, results about centrally large subgroups of p-groups. A subgroup H of a finite p-group P is *centrally large* if $|H||Z(H)| \geq |H^*||Z(H^*)|$ for every subgroup H^* of P.

For any positive real number α, Chermak and Delgado defined the measure

$$m_\alpha(G, H) = \mathrm{Sup}\{|A|^\alpha |C_H(A)|\}_{A \in S(G)},$$

where $S(G)$ is the set of all nontrivial subgroups of G. In his book, I. M. Isaacs [2008] focused on the case where $\alpha = 1$ and H is a subgroup of the finite group G. He defined the *Chermak–Delgado measure* of a subgroup H of G as $m_G(H) = |H||C_G(H)|$. In a definition analogous to Glauberman's centrally large subgroups, Isaacs defined a subgroup H of G to have *maximal Chermak–Delgado measure* if $|H||C_G(H)| \geq |H^*||C_G(H^*)|$ for every subgroup H^* of G. Finally, Isaacs showed that the subgroups with maximal Chermak–Delgado measure form a sublattice of the lattice of all subgroups of G, which he termed the *Chermak–Delgado lattice of* G.

MSC2010: 20D30.

Keywords: centrally large subgroups, Chermak–Delgado measure, lattices of subgroups, metacyclic p-groups.

Not only are the definitions similar, but one can see that H is centrally large in a p-group P if and only if H is in the Chermak–Delgado lattice of P and $C_P(H) \leq H$ [Glauberman 2006]. Thus, the study of Chermak–Delgado lattices may shed light on the subgroup structure of finite groups in general, and on centrally large subgroups in particular.

Not much is known about Chermak–Delgado lattices. To date, there are only four published papers on the subject. B. Brewster and E. Wilcox [2012] studied the connection between Chermak–Delgado lattices of direct and wreath products of groups and the Chermak–Delgado lattices of their components. Together with P. Hauck they constructed a class of p-groups whose Chermak–Delgado lattice is a chain [Brewster et al. 2014a]. Later the trio studied quasiantichains in Chermak–Delgado lattices, proving that if there is a quasiantichain interval between subgroups L and H of G, with $L \leq H$, then H/L is an elementary abelian p-group for some prime p [Brewster et al. 2014b]. The most recent contribution to the subject is a paper by L. An, J. Brennan, H. Qu, and Wilcox [An et al. 2015], who further studied groups for which the Chermak–Delgado lattice is a chain of diamonds.

The structure of a Chermak–Delgado lattice varies greatly among finite groups. For example, it is easy to see that the Chermak–Delgado lattice of an abelian group G consists of just G. There are groups whose Chermak–Delgado lattices are chains of any length, diamonds of any width, and chains of diamonds; see [Brewster et al. 2014a; An et al. 2015]. These two papers construct groups that have particular Chermak–Delgado lattice shapes, whereas we start with general split metacyclic p-groups, $p > 2$, and construct their associated Chermak–Delgado lattices.

The structure of this paper is as follows. In Section 2, we define some terms and describe the symmetry of Chermak–Delgado lattices in general. In Section 3, we give a full treatment of split metacyclic p-groups, $p > 2$, from describing their subgroup structure to determining their Chermak–Delgado measure and lattices. In Section 4, we define a rank-symmetric subposet of the poset of all subgroups of a split metacyclic p-group with maximal Chermak–Delgado measure; sometimes its corresponding lattice has the same structure as the full Chermak–Delgado lattice. Finally, in Section 5, we construct two complete Chermak–Delgado lattices using the theorems developed in Section 3.

2. Chermak–Delgado lattices and their symmetry

A good reference for Chermak–Delgado measures and lattices is [Isaacs 2008], where we find some of the definitions and properties below.

As above, if G is a finite group with subgroup H, then the *Chermak–Delgado measure* (or *CD-measure*) of H in G is $m_G(H) = |H||C_G(H)|$, where $C_G(H)$ is the centralizer of H in G. We will denote the maximum possible CD-measure in a group G as $m^*(G)$. The set of all subgroups of G that have maximal measure is a poset called the *Chermak–Delgado set* of G (or *CD-set*) and denoted $\mathcal{CD}(G)$.

As with any poset, we are interested in smallest and largest elements. The greatest lower bound for the entirety of $\mathcal{CD}(G)$ is the intersection of all members of the set, which is called the *Chermak–Delgado subgroup* of G and denoted M_G. Isaacs proved that $Z(G) \leq M_G$. It follows that $Z(G) \leq H$ for any $H \in \mathcal{CD}(G)$. Isaacs further proved that if $H \in \mathcal{CD}(G)$, then $C_G(C_G(H)) = H$; hence $C_G(H)$ is also in $\mathcal{CD}(G)$. In particular, $M^G = C_G(M_G)$ is the least upper bound for the entirety of $\mathcal{CD}(G)$ and $m^*(G) = |M_G||M^G|$.

The *Chermak–Delgado lattice* of G (or *CD-lattice*) consisting of subgroups from $\mathcal{CD}(G)$ is a modular, self-dual lattice [Brewster et al. 2014b]. Usually this lattice is also denoted by $\mathcal{CD}(G)$, but we will use the notation $\mathcal{L}(G)$ to specifically denote the Hasse diagram drawn from the CD-set such that there is an edge between subgroups $H_1, H_2 \in \mathcal{CD}(G)$ if and only if H_1 *covers* H_2, meaning that $H_2 < H_1$ and there does not exist a subgroup $K \in \mathcal{CD}(G)$ for which $H_2 < K < H_1$. Although subgroup order does not necessarily define a rank function on $\mathcal{CD}(G)$, we will see that the lattice $\mathcal{L}(G)$ has several interesting properties.

Define the *height* of a subgroup H in the lattice $\mathcal{L}(G)$ as simply $|H|$. If H_1 and H_2 are two subgroups in $\mathcal{L}(G)$ of orders n_1 and n_2 respectively, with $n_1 \geq n_2$, we will call n_1/n_2 the *distance* between H_1 and H_2 and denote it $d(H_1, H_2)$. In our scheme, we have $d(H, K) \geq 1$ for all $H, K \in \mathcal{CD}(G)$.

The next theorem says $\mathcal{CD}(G)$ is order symmetric and $\mathcal{L}(G)$ can be displayed having a horizontal line of symmetry.

Theorem 2.1. *Let G be a finite group. Then $\mathcal{L}(G)$ is graph isomorphic to a lattice that is symmetric across a horizontal line of symmetry at height $\sqrt{m^*(G)}$.*

Proof. To prove the symmetry, we will show that the bijective correspondence $H \mapsto C_G(H)$, for $H \in \mathcal{CD}(G)$, satisfies three properties:

(1) If H lies above the line of symmetry, then the distance between the top of the lattice and H is equal to the distance between $C_G(H)$ and the bottom of the lattice (by duality, the analogous property will hold for any H lying below the line of symmetry).

(2) If H lies on the line of symmetry, then so does $C_G(H)$.

(3) Subgroup inclusion relationships above the line of symmetry are mirrored below the line.

Thus, a symmetric lattice can be drawn with $H \in \mathcal{L}(G)$ and its partner $C_G(H)$ placed on a vertical line above and below the line of symmetry, or both exactly on the line. (Figure 1 shows the same portion of the CD-lattice for the metacyclic p-group $P(4, 2, 2, 0)$ as described in (3-1) below. The lattice on the right is drawn

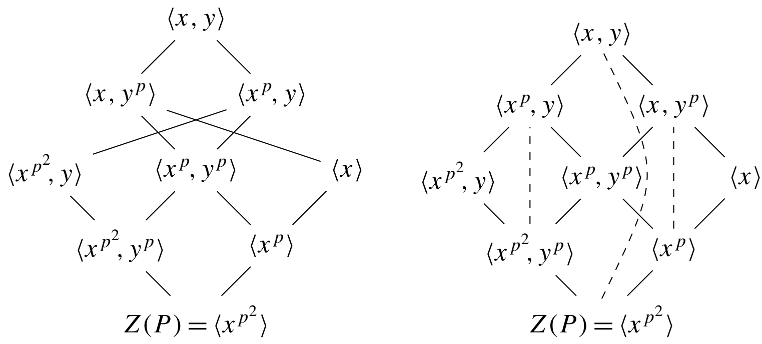


Figure 1. Isomorphic portions of the CD-lattice for $P(4, 2, 2, 0)$.

with its symmetry highlighted, and dashed lines connecting a subgroup to its centralizer.)

Proof of (1): Let $H \in \mathcal{CD}(G)$ and assume $|H| > \sqrt{m^*(G)}$. Note that

$$|H||C_G(H)| = m^*(G) = |M_G||M^G|.$$

It follows that

$$\frac{|C_G(H)|}{|M_G|} = \frac{|M^G|}{|H|}.$$

Hence, $d(C_G(H), M_G) = d(M^G, H)$.

Proof of (2): Assume $H \in \mathcal{CD}(G)$ lies on the line of symmetry, so $|H| = \sqrt{m^*(G)}$. Since $|H||C_G(H)| = m^*(G)$, we must have $|C_G(H)| = \sqrt{m^*(G)}$, so $C_G(H)$ also lies on the line of symmetry.

Proof of (3): Assume $H, K \in \mathcal{CD}(G)$ and $H \leq K$. It follows that $C_G(K) \leq C_G(H)$. Thus, subgroup inclusion relationships above the line of symmetry are mirrored below the line, and vice versa. □

3. Subgroup and CD-lattices of split metacyclic p-groups

Presentations of metacyclic p-groups, p > 2. A group is metacyclic if it has a cyclic normal subgroup whose corresponding quotient is also cyclic. The group splits if it is a semidirect product of the form $\mathbb{Z}_m \rtimes \mathbb{Z}_n$. It is well known that every noncyclic metacyclic p-group, $p > 2$, has a presentation of the form

$$P = P(m, n, c, s) = \langle x, y \mid x^{p^m} = 1,\ y^{p^n} = x^{p^{m-s}},\ yxy^{-1} = x^{1+p^{m-c}}\rangle, \qquad (3\text{-}1)$$

where $1 \leq m, n,\ 0 \leq c \leq \min\{m-1, n\}$, and $0 \leq s \leq m - c$.[1] Note that the parameter s measures how far the group is from splitting (if $s = 0$, then P is split

[1]Presentations of metacyclic 2-groups are also known, but we restrict to odd primes throughout.

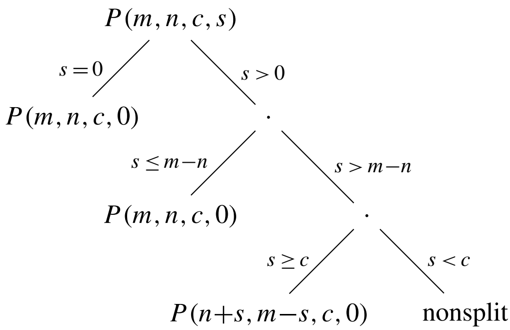


Figure 2. Conditions under which $P(m, n, c, s)$ splits.

and $P \cong \langle x \rangle \rtimes \langle y \rangle$), and the parameter c measures how far the group is from being commutative (if $c = 0$, then P is abelian). We have $|P| = p^{m+n}$ and every element of P can be written uniquely as $x^i y^j$ for some $0 \leq i < p^m$ and $0 \leq j < p^n$.

Note 3.1. For the rest of this paper, we will use the term "metacyclic p-group" to mean a finite, nonabelian, noncyclic metacyclic p-group, where $p > 2$.

Depending on the relative sizes of the four parameters, it is sometimes possible to find a different set of generators that yields a split presentation with $s = 0$.

Proposition 3.2 [Dietz 1993]. *Let $P(m, n, c, s)$ be as in (3-1).*

(1) *If $s = 0$, then P is split.*

(2) *If $s \neq 0$ and $m - s \geq n$, then we can find an element y^* such that $P \cong \langle x, y^* \rangle = P(m, n, c, 0)$ and thus P splits.*

(3) *If $s \neq 0$ and $m - s < \min\{n, m - c + 1\}$, then we can find elements x^* and y^* such that $P \cong \langle y^*, x^* \rangle = P(n + s, m - s, c, 0)$ and thus P splits.*

(4) *If $s \neq 0$ and $m - c < m - s < n$, then P is nonsplit.*

See Figure 2 for a decision tree illustrating this proposition.

Furthermore, we can put restrictions on the parameters in such a way that different values of s and c for a fixed pair m and n describe unique isomorphism types. We have the following isomorphism classes that we will say are in *reduced* metacyclic form.

Proposition 3.3 [King 1973]. *Every metacyclic p-group is isomorphic to exactly one of the following reduced forms*:

(1) *Split*: $P(m, n, c, s)$ *with* $0 = s \leq c \leq \min\{n, m - 1\}$.

(2) *Nonsplit*: $P(m, n, c, s)$ *with* $\max\{1, m - n + 1\} \leq s \leq \min\{c - 1, m - c\}$.

CD-measure and CD-lattice of $P(m, n, c, s)$. We first establish the CD-measures of split metacyclic p-groups, and then determine their CD-lattices.

Proposition 3.4. *Let $P = P(m, n, c, 0)$ be a split metacyclic p-group. Then $m^*(P) = p^{2(m+n-c)}$.*

Proof. By [Héthelyi and Külshammer 2011, Lemma 4.1], we have $m^*(P) = [P : P']^2$. Lemma 2.5 in [Bidwell and Curran 2010] shows that $P' = \langle x^{p^{m-c}} \rangle \cong \mathbb{Z}_{p^c}$. Hence $m^*(P) = (p^{m+n-c})^2$. □

Theorem 3.5. *Let $P = P(m, n, c, 0)$. Then $\mathcal{CD}(P)$ consists of the maximal abelian subgroups of P, all nonabelian subgroups of P, and all centralizers of the nonabelian subgroups.*

Proof. First, we show that $\mathcal{CD}(P)$ contains the subgroups mentioned in the theorem; second, we show that no other subgroups are in $\mathcal{CD}(P)$.

Proposition 2.4 in [Glauberman 2006] shows that all centrally large subgroups of a finite p-group have maximal CD-measure. Corollary 4.4 in [Héthelyi and Külshammer 2011] shows that all maximal abelian and nonabelian subgroups of P are centrally large; hence all are in $\mathcal{CD}(P)$. By duality, centralizers of these subgroups are also in $\mathcal{CD}(P)$. We will show that the maximal abelian subgroups are equal to their centralizers.

Let A be a maximal abelian subgroup of P. By Lemma 2.8 in [Héthelyi and Külshammer 2011], $|A| = [P : P'] = p^{m+n-c}$. Since A has maximal CD-measure, we know $|C_P(A)| = p^{m+n-c}$. Finally, $A \le C_P(A)$ implies $A = C_P(A)$.

Next, we must show that no other subgroups of P are contained in $\mathcal{CD}(P)$. Let H be an abelian subgroup of P that is not maximal abelian and is not the centralizer of a nonabelian subgroup of P. If $C_P(H)$ is nonabelian, then $C_P(H) \in \mathcal{CD}(P)$. By duality, $C_P(C_P(H)) = H$, but this contradicts the fact that H is not the centralizer of a nonabelian subgroup of P. If $C_P(H)$ is abelian, then neither H nor $C_P(H)$ is maximal abelian. We see that

$$m_P(H) = |H||C_P(H)| < (p^{m+n-c})^2 \le m^*(P),$$

so $H \notin \mathcal{CD}(P)$. □

A few things from the theorem and its proof are worthy of note.

- P is itself a member of $\mathcal{CD}(P)$, hence $Z(P)$ is too.
- By Theorem 2.1, the line of symmetry of $\mathcal{L}(P)$ is at height $\sqrt{m^*(P)} = p^{m+n-c}$, which is the height of the maximal abelian subgroups of P. Thus, it makes sense that the maximal abelian subgroups of P are equal to their own centralizers.
- All nonabelian subgroups of P lie above the line of symmetry in $\mathcal{L}(P)$ and have order greater than all abelian subgroups of P (this is in Proposition 3.3 of [Berkovich 2013] too).

• All nonabelian subgroups of P have abelian centralizers (that are not maximal). In fact, the proposition below shows that centralizers of nonabelian subgroups equal their centers.

• All nonabelian subgroups of P contain the center of P.

We have two further results about abelian and nonabelian subgroups of $P = P(m, n, c, 0)$.

Proposition 3.6. *Let H be a nonabelian subgroup of P. Then $C_P(H) = Z(H)$.*

Proof. As we saw in the proof above, H is centrally large, so $|H||Z(H)| \geq |H^*||Z(H^*)|$ for all $H^* \leq P$. Since a maximal abelian subgroup A of P is equal to its own centralizer, we see that $|A||C_P(A)| = |A||Z(A)|$ is maximal too. Then

$$|H||Z(H)| \geq |A||Z(A)| = |A||C_P(A)| = |H||C_P(H)|$$

implies $C_P(H) = Z(H)$. □

Proposition 3.7. *Not every nonmaximal abelian subgroup A of P with $|A| \geq |Z(P)|$ is in $\mathcal{CD}(P)$.*

Proof. First some notation: let s_j denote the total number of subgroups of P of order p^j, and let s_j^{cd} denote the total number of subgroups of order p^j that are in $\mathcal{CD}(P)$.

From Theorem 3.5 we know that if a nonmaximal abelian subgroup of P is in $\mathcal{CD}(P)$, then it has order p^k, where $m+n-2c \leq k < m+n-c$. Thus, if we can show $s_k > s_k^{\mathrm{cd}}$ for all such k, then we will have proved the proposition.

By duality, $s_k^{\mathrm{cd}} = s_i^{\mathrm{cd}}$, where $i = 2(m+n-c)-k$. Since every nonabelian subgroup of P is in $\mathcal{CD}(P)$, we know $s_i^{\mathrm{cd}} = s_i$. By a result of A. Mann [2010], the number of subgroups of P of order p^i, where $m+n-c < i \leq m+n$, is

$$s_i = \frac{p^{m+n-i+1}-1}{p-1}.$$

The same theorem in [Mann 2010] shows that the size of s_k depends on the size of k relative to m and n.

Assume $m \geq n$.

(1) If $k \geq m$, then $s_k = (p^{m+n-k+1}-1)/(p-1)$. Since $k < i$, we have $s_k > s_i$.

(2) If $n \leq k \leq m$, then $s_k = (p^{n+1}-1)/(p-1)$. Since $m+n-c < i \leq m+n$ and $c \leq n$, we have $i > m$. Thus $n+1 > m+n-i+1$ and $s_k > s_i$.

(3) If $k \leq n$, then $s_k = (p^{k+1}-1)/(p-1)$. Since $c \leq \min\{m-1, n\}$, we have $m+n-2c > 0$. Thus

$$k+1 = 2(m+n-c)-i+1 > m+n-i+1,$$

and $s_k > s_i$.

In each case, the total number of abelian subgroups of order p^k is greater than the number of abelian subgroups of order p^k that appear in $\mathcal{CD}(P)$.

A similar result holds if $m < n$. $\square$

By duality, we need only determine the upper half of the CD-lattice in order to at least know its complete structure (if not the precise subgroups). The next subsection gives us tools for determining the nonabelian subgroups of $P(m, n, c, 0)$ in a regressive manner.

Subgroups of metacyclic p-groups. It is well known that subgroups of metacyclic groups are either cyclic or metacyclic. In this section we determine the precise metacyclic structure of the maximal subgroups of $P = P(m, n, c, s)$. First we need some computational lemmas from [Schulte 2001] and [Bidwell and Curran 2010].[2]

Lemma 3.8 [Schulte 2001]. *Let $P = P(m, n, c, s)$ be as in (3-1).*

(1) *We have $y^j x^i = x^{i(1+p^{m-c})^j} y^j$, where $i, j \geq 0$.*

(2) *Let $\alpha = 1 + p^{m-c}$ and $k \in \mathbb{N}$. Then $(x^i y^j)^k = x^{i\Lambda(j,k)} y^{jk}$, where*

$$\Lambda(j, k) = 1 + \alpha^j + \alpha^{2j} + \cdots + \alpha^{(k-1)j} \quad \textit{for } i, j, k \geq 0.$$

Lemma 3.9 [Bidwell and Curran 2010]. *For any positive integers a and b and an odd prime p, we have $(1 + p^a)^{p^b} \equiv 1 \bmod p^{a+b}$.*

From the proof of Proposition 3.7, the number of proper subgroups of P of maximal order is $s_i = p + 1$, where $i = m + n - 1$.

Theorem 3.10. *The $p+1$ maximal proper subgroups of $P = P(m, n, c, s) = \langle x, y\rangle$ are*

$$L = \langle x^p, y\rangle, \qquad M_i = \langle x^p, x^i y\rangle \quad \textit{for } i = 1, \ldots, p-1, \qquad R = \langle x, y^p\rangle.$$

Proof. It is clear that L and R each have order p^{m+n-1} and are distinct from one another. L and M_i are distinct because $y \in M_i$ only if $i \equiv 0 \bmod p$. Similarly, R and M_i are distinct because $x \notin M_i$. To show the M_i are distinct from one another, suppose that $x^i y \in M_j$ for some $i \neq j$. Then for some $u, v \in \mathbb{Z}$, we have

$$x^i y = (x^p)^u (x^j y)^v = x^{pu + j\Lambda(1,v)} y^v$$

by Lemma 3.8. We must have $v \equiv 1 \bmod p^n$ so there exists $w \in \mathbb{Z}$ such that $v = 1 + wp^n$. Now

$$x^i y = x^{pu + j\Lambda(1,v)} y^{1+wp^n} = x^{pu + j\Lambda(1,v) + wp^{m-s}} y,$$

hence $i \equiv pu + j\Lambda(1, v) + wp^{m-s} \bmod p^m$. By Lemma 3.9 we can see that $\Lambda(1, v) \equiv 1 \bmod p$. Thus $i \equiv j \bmod p$, which is a contradiction.

[2]Schulte [2001] focuses on a particular family of split metacyclic p-groups, but the result stated here clearly holds more generally.

It remains to show that $|M_i| = p^{m+n-1}$. Because the M_i are metacyclic, we know that

$$|M_i| = \frac{|\langle x^p\rangle||\langle x^i y\rangle|}{|\langle x^p\rangle \cap \langle x^i y\rangle|}.$$

Using Lemmas 3.8 and 3.9, we can show that $|x^i y| \geq |y| = p^{n+s}$. Using similar computations, we can show that $\langle x^p\rangle \cap \langle x^i y\rangle \leq \langle x^{p^{m-s}}\rangle$. Thus

$$\frac{|\langle x^p\rangle||\langle x^i y\rangle|}{|\langle x^p\rangle \cap \langle x^i y\rangle|} \geq \frac{p^{m-1}p^{n+s}}{p^s} = p^{m+n-1}.$$

Certainly M_i is a proper subgroup of P, so we must have $|M_i| = p^{m+n-1}$. □

Next we identify the metacyclic structure of the $p+1$ maximal subgroups of a metacyclic p-group, but separate the split and nonsplit cases.

Theorem 3.11. *The metacyclic forms of the maximal subgroups of the split metacyclic p-group $P = P(m, n, c, 0)$ are*

(1) $L = \langle x^p, y\rangle \cong P(m-1, n, c-1, 0)$,

(2) $M_i = \langle x^p, x^i y\rangle \cong \begin{cases} P(m-1, n, c-1, 0) \cong L & \text{if } m \leq n, \\ P(m, n-1, c-1, 0) \cong R & \text{if } m > n \text{ and } n \leq m-c+1, \\ P(m-1, n, c-1, m-n) & \text{if } m > n > m-c+1, \end{cases}$

(3) $R = \langle x, y^p\rangle \cong P(m, n-1, c-1, 0)$.

Proof. The order relations in L are clear, so we need only check the degree of commutativity. We have

$$yx^p y^{-1} = (yxy^{-1})^p = (x^{1+p^{m-c}})^p = (x^p)^{1+p^{m-c}}.$$

Since $m-c = (m-1)-(c-1)$, we have our result.

In R we see that $y^p x y^{-p} = x^{(1+p^{m-c})^p}$ by Lemma 3.8. By Lemma 3.9 we know that $(1+p^{m-c})^p \equiv 1 \bmod p^{m-c+1}$. By [King 1973, Proposition 2.3], we can replace x and y^p with x^* and y^* respectively so that $\langle x^*\rangle = \langle x\rangle$, $\langle y^*\rangle = \langle y^p\rangle$ and $y^* x^* (y^*)^{-1} = (x^*)^{1+p^{m-c+1}}$. Since $m-c+1 = m-(c-1)$, we have our result.

In M_i we have

$$(x^i y)x^p(x^i y)^{-1} = x^i(yx^p y^{-1})x^{-i} = x^i(x^{1+p^{m-c}})^p x^{-i} = (x^p)^{1+p^{m-c}}.$$

Next, we compute the splitting degree. Since P is regular (see [Davitt 1970, Corollary 1]), we know

$$(x^i y)^{p^n} = x^{ip^n} y^{p^n} z^{p^n}$$

for some $z \in [P, P] = \langle x^{p^{m-c}}\rangle$. Suppose $z = x^{ap^{m-c}}$ for some $a \in \mathbb{Z}$. Then

$$(x^i y)^{p^n} = x^{ip^n + ap^{m-c+n}}.$$

If $m \le n$, we see that $x^i y$ has order p^n and $M_i = P(m-1,\, n,\, c-1,\, 0)$. If $m > n$, then

$$(x^i y)^{p^n} = (x^p)^{ip^{n-1} + ap^{m-c+n-1}} = (x^p)^{p^{n-1}(i+ai^{-1}p^{m-c})}.$$

Since $i + ai^{-1}p^{m-c}$ is a unit modulo p^m, [King 1973] again tells us that we can choose new generators x^* and y^* so that $\langle x^*\rangle = \langle x^p\rangle$, $\langle y^*\rangle = \langle x^i y\rangle$, $y^* x^* (y^*)^{-1} = (x^*)^{1+p^{m-c}}$, and $(y^*)^{p^n} = (x^*)^{p^{n-1}}$. Hence we get $M_i = P(m-1,\, n,\, c-1,\, m-n)$. Finally, by Proposition 3.2 we see that the reduced metacyclic form of M_i depends on whether $m - n < c - 1$. □

Theorem 3.12. *The metacyclic forms of the maximal subgroups of the nonsplit metacyclic p-group* $P = P(m, n, c, s)$, $\max\{1, m-n+1\} \le s \le \min\{c-1, m-c\}$, *are*

(1) $L = \langle x^p, y\rangle \cong \begin{cases} P(n+s,\, m-s-1,\, c-1,\, 0) & \textit{if } s = c-1, \\ P(m-1,\, n,\, c-1,\, s) & \textit{if } s < c-1, \end{cases}$

(2) $M_i = \langle x^p, x^i y\rangle \cong L$ *for* $i = 1, \dots, p-1$,

(3) $R = \langle x, y^p\rangle \cong \begin{cases} P(m,\, n-1,\, c-1,\, 0) & \textit{if } s = m-n+1, \\ P(n+s-1,\, m-s,\, c-1,\, 0) & \textit{if } m-n+1 < s = c-1, \\ P(m,\, n-1,\, c-1,\, s) & \textit{if } m-n+1 < s < c-1. \end{cases}$

Proof. Consider the subgroup L first. As in the proof of Theorem 3.11, we have the commutativity degree is $c-1$. Since $y^{p^n} = (x^p)^{p^{m-1-s}}$, we see that the splitting degree is s. Thus $L = P(m-1,\, n,\, c-1,\, s)$. By Proposition 3.2 and the subsequent decision tree, we know $s > 0$ and $s \ge m-n+1 > (m-1)-n$, so that the structure of L depends on whether $s \ge c-1$. We already have $s \le c-1$, so $s \ge c-1$ if and only if $s = c-1$.

Next we consider R. As in the proof of Theorem 3.11, we can replace x and y^p with x^* and y^* respectively so that $y^* x^* (y^*)^{-1} = (x^*)^{1+p^{m-c+1}}$. Furthermore, the splitting degree remains unchanged, so $(y^p)^{p^{n-1}} = x^{p^{m-s}}$ implies $(y^*)^{p^{n-1}} = (x^*)^{p^{m-s}}$. Hence $R = P(m,\, n-1,\, c-1,\, s)$. From the decision tree, we see that $R = P(m,\, n-1,\, c-1,\, 0)$ if $s \le m-(n-1)$. We already have $m-n+1 \le s$, so this form occurs exactly when $s = m-n+1$. Finally, the structure of R depends on whether $s \ge c-1$ (which happens if and only if $s = c-1$) or $s < c-1$.

Lastly, we consider the subgroups M_i. We compute $x^i y x^p y^{-1} x^{-p} = (x^p)^{1+p^{m-c}}$ so the commutativity degree is $c-1$. The splitting degree is determined by computing $(x^i y)^{p^n} = x^{i\Lambda(1,p^n)+p^{m-s}}$, where $\Lambda(1, p^n)$ is as in Lemma 3.8. Now

$$\Lambda(1, p^n) = 1 + \alpha + \alpha^2 + \cdots + \alpha^{p^n-1} = \frac{\alpha^{p^n} - 1}{\alpha - 1}.$$

Since $\alpha^{p^n} \equiv 1 \bmod p^{m-c+n}$ by Lemma 3.9, there exists $a \in \mathbb{Z}$ such that $\alpha^{p^n} = 1 + ap^{m-c+n}$. Thus $\Lambda(1, p^n) = ap^n$ and we have

$$(x^i y)^{p^n} = x^{iap^n + p^{m-s}} = x^{p^{m-s}(iap^{n-m+s}+1)} = (x^p)^{p^{m-s-1}(iap^{n-m+s}+1)}.$$

Once again by [King 1973] we can replace x^p and $x^i y$ with generators x^* and y^* satisfying the same order and commutator relations, and further satisfying $(y^*)^{p^n} = (x^*)^{p^{m-1-s}}$. Hence $M_i = P(m-1, n, c-1, s)$ and has the same structure as L. □

With Theorems 3.5, 3.11, and 3.12 in hand, we can construct the subgroup lattice of a split metacyclic p-group, and hence construct its CD-lattice. In practice, even small-order metacyclic p-groups will have complicated lattices with split subgroups spawning nonsplit subgroups, and vice versa (see one example in Section 5). However, there are certain conditions under which we get particularly nice CD-lattices, including some quasiantichains as discussed in [Brewster et al. 2014b]. This is what we will describe in the next section.

4. Diamonds in the rough: the BEK-lattice

We begin with some notation. Let $P = P(m, n, c, 0) = \langle x, y\rangle$, and set $H_{ab} = \langle x^{p^a}, y^{p^b}\rangle$, where $0 \le a \le m$ and $0 \le b \le n$. These subgroups are the L- and R-types from Theorems 3.11 and 3.12, and sometimes the M_i subgroups are isomorphic to these. The set of the H_{ab} will form a sublattice of $\mathcal{L}(P)$, which we will denote $\mathcal{BEK}(P)$, and under the right circumstances $\mathcal{L}(P)$ will "collapse" to $\mathcal{BEK}(P)$. Before getting to these results, we delve into the structure of the H_{ab}.

Lemma 4.1 [Bidwell and Curran 2010]. *Let $P = P(m, n, c, s)$ be as in (3-1). Then*

(1) $C_{\langle x\rangle}(\langle y\rangle) = \langle x^{p^c}\rangle$,

(2) $C_{\langle y\rangle}(\langle x\rangle) = \langle y^{p^c}\rangle$,

(3) $Z(P) = \langle x^{p^c}, y^{p^c}\rangle$ *and* $|Z(P)| = p^{m+n-2c}$.

By the third property above, we have $Z(P) \le H_{ab}$ for all $0 \le a, b \le c$.

We know the metacyclic forms of the subgroups H_{ab}.

Proposition 4.2. *Let $P = P(m, n, c, 0)$ and consider $H_{ab} \le P$, where $0 \le a \le m$ and $0 \le b \le n$. Then*

$$H_{ab} \cong \begin{cases} P(m-a,\, n-b,\, c-(a+b),\, 0) & \text{if } a+b < c,\\ \mathbb{Z}_{p^{m-a}} \times \mathbb{Z}_{p^{n-b}} & \text{if } a+b \ge c.\end{cases}$$

In particular, $|H_{ab}| = p^{m+n-a-b}$.

Proof. By repeated applications of Theorem 3.11, we see that

$$\langle x^{p^a}, y^{p^b}\rangle = P(m-a,\, n-b,\, c-(a+b),\, 0)$$

as long as $c-(a+b) > 0$. On the other hand, by Lemma 3.8

$$y^{p^b} x^{p^a} = x^{p^a(1+p^{m-c})^{p^b}} y^{p^b}.$$

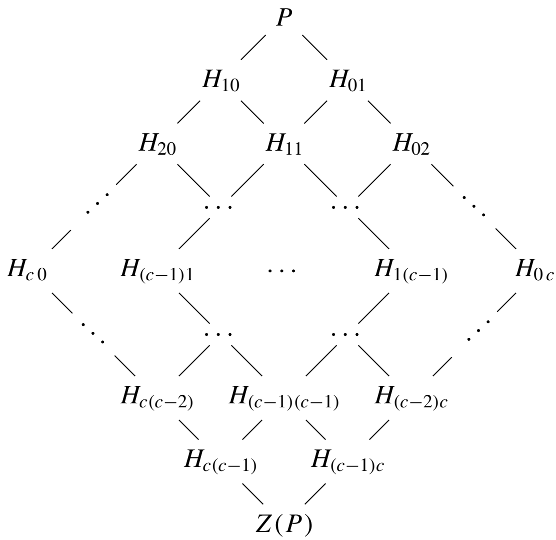


Figure 3. The BEK-lattice of a split metacyclic p-group.

By Lemma 3.9, $(1+p^{m-c})^{p^b} \equiv 1 \mod p^{m-c+b}$. Thus

$$x^{p^a(1+p^{m-c})^{p^b}} = x^{p^a(1+dp^{m-c+b})}$$

for some $d \in \mathbb{Z}$. If $a+b \geq c$, we can see that H_{ab} is abelian and isomorphic to $\mathbb{Z}_{p^{m-a}} \times \mathbb{Z}_{p^{n-b}}$. □

Theorem 4.3. *Let $\mathcal{CD}_{\text{bek}}(P) = \{H_{ab} \mid 0 \leq a, b \leq c\}$. Then $\mathcal{CD}_{\text{bek}}(P)$ is a subposet of $\mathcal{CD}(P)$ that is rank-symmetric, and its corresponding lattice, $\mathcal{BEK}(P)$, is shown in Figure 3.*

To show that each H_{ab} is in $\mathcal{CD}(P)$, we will show that the centralizers of the nonabelian H_{ab} are the abelian ones.

Proposition 4.4. $$C_P(H_{ab}) = H_{c-b,c-a}.$$

Proof. When $a+b < c$, we know H_{ab} is nonabelian and thus is in $\mathcal{CD}(P)$ by Theorem 3.5. Therefore, $m_P(H_{ab}) = p^{2(m+n-c)}$. Since $|H_{ab}| = p^{m+n-a-b}$, it follows that $|C_P(H_{ab})| = p^{m+n-2c+a+b}$. By Lemma 4.1,

$$Z(H_{ab}) = Z(P(m-a,\, n-b,\, c-a-b,\, 0)) = \langle x^{p^{c-b}}, y^{p^{c-a}} \rangle = H_{c-b,c-a}$$

and therefore $|Z(H_{ab})| = p^{m+n-2c+a+b}$. We have $|C_P(H_{ab})| = |Z(H_{ab})|$ and so $C_P(H_{ab}) = Z(H_{ab}) = H_{c-b,c-a}$.

When $a+b = c$, we know $|H_{ab}| = p^{m+n-c}$, so it is maximal abelian. From Theorem 3.5 we know $C_P(H_{ab}) = H_{ab} = H_{c-b,c-a}$.

Let $a+b>c$. Then by Lemma 3.8,

$$y^{p^b}x^{p^{c-b}} = x^{p^{c-b}(1+p^{m-c})^{p^b}}y^{p^b} = x^{p^{c-b}}y^{p^b},$$

so $x^{p^{c-b}} \in C_P(H_{ab})$. Similarly,

$$y^{p^{c-a}}x^{p^a} = x^{p^a(1+p^{m-c})^{p^{c-a}}}y^{p^{c-a}} = x^{p^a}y^{p^{c-a}},$$

so $y^{p^{c-a}} \in C_P(H_{ab})$. Thus $H_{c-b,c-a} \le C_P(H_{ab})$. We have

$$|H_{ab}|\,|C_P(H_{ab})| \ge |H_{ab}|\,|H_{c-b,c-a}| = p^{2m+2n-2c}.$$

Since this is the maximal possible measure, we have $H_{c-b,c-a} = C_P(H_{ab})$. □

The next corollary follows immediately.

Corollary 4.5. $H_{ab} \in \mathcal{CD}(P)$ *for all* $0 \le a, b \le c$.

Proposition 4.6. *H_{ab} covers $H_{a'b'}$ if and only if $a+b+1 = a'+b'$ and either $a' = a+1$ or $b' = b+1$.*

Proof. First suppose $H_{a'b'} < H_{ab}$. The generators of $H_{a'b'}$ must be in H_{ab}, so there exist $i, j, u, v \in \mathbb{Z}$ such that

$$x^{p^{a'}} = (x^{p^a})^i(y^{p^b})^j, \tag{4-1}$$

$$y^{p^{b'}} = (x^{p^a})^u(y^{p^b})^v. \tag{4-2}$$

From (4-1) we see that $p^{a'} = ip^a$, so $a' \ge a$. From (4-2) we have $b' \ge b$. If $a+b = a'+b'-1$, then there are exactly two groups that cover $H_{a'b'}$ as stated in the proposition. If $a+b < a'+b'-1$, then one of three cases can occur: (i) $a \le a'-2$, (ii) $b \le b'-2$, or (iii) $a \le a'-1$ and $b \le b'-1$. In turn, we will have the three cases

(i) $H_{a'b'} < H_{a'-1,b'} < H_{a'-2,b'} \le H_{ab}$,

(ii) $H_{a'b'} < H_{a',b'-1} < H_{a',b'-2} \le H_{ab}$,

(iii) $H_{a'b'} < H_{a'-1,b'} < H_{a'-1,b'-1} \le H_{ab}$,

contradicting the fact that H_{ab} covers $H_{a'b'}$.

Conversely, if $a+b+1 = a'+b'$ and $a' = a+1$, then it is clear that $H_{a'b'} < H_{ab}$ and there does not exist H_{cd} such that $H_{a'b'} < H_{cd} < H_{ab}$. We have a similar result if $b' = b+1$. □

The proposition above proves that $\mathcal{BEK}(P)$ has the structure illustrated in Figure 3.

Finally, we define a rank function $\rho : \mathcal{CD}_{\text{bek}}(P) \to \mathbb{N}$ by $\rho(H_{ab}) = m+n-a-b$. Then $\rho(H_{ab}) = \rho(H_{a'b'})+1$ when H_{ab} covers $H_{a'b'}$ and we see that $\mathcal{CD}_{\text{bek}}(P)$ is a ranked poset that is clearly rank-symmetric from its definition, proving Theorem 4.3.

$$
\begin{tikzcd}[column sep=tiny]
 & & & P \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & & & \\
 & & H_{10}(p) \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & & H_{01}(1) \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & & \\
 & H_{20}(p^2) \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & & H_{11}(p) \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & & H_{02}(1) \arrow[dl, dash, "p"'] \arrow[dr, dash, "1"] & \\
H_{30}(p^3) \arrow[dr, dash, "1"'] & & H_{21}(p^2) \arrow[dl, dash, "p"] \arrow[dr, dash, "1"'] & & H_{12}(p) \arrow[dl, dash, "p"] \arrow[dr, dash, "1"'] & & H_{03}(1) \arrow[dl, dash, "1"] \\
 & H_{31}(p^2) \arrow[dr, dash, "1"'] & & H_{22}(p) \arrow[dl, dash, "p"] \arrow[dr, dash, "1"'] & & H_{13}(1) \arrow[dl, dash, "p"] & \\
 & & H_{32}(p) \arrow[dr, dash, "1"'] & & H_{23}(1) \arrow[dl, dash, "p"] & & \\
 & & & Z(P) & & &
\end{tikzcd}
$$

Figure 4. The CD-lattice of $P(4, 6, 3, 0)$.

As we will prove below, there are conditions on the parameters of $P(m, n, c, 0)$ that guarantee its CD-lattice "collapses" to the BEK-lattice. That is, all subgroups in $\mathcal{CD}(P)$ will be isomorphic to H_{ab} for some $0 \le a, b \le c$. There will be multiple copies of some H_{ab} in $\mathcal{CD}(P)$ — for example, in $P(3, 4, 1, 0)$ there will be p copies of H_{10} by Theorem 3.11 — so that the edges in $\mathcal{L}(P)$ will be weighted versions of those in $\mathcal{BEK}(P)$.

Theorem 4.7. *Let $P = P(m, n, c, 0)$, with $m \le n$ and $c \le n - m + 1$. Then every subgroup in $\mathcal{CD}(P)$ is isomorphic to H_{ab} for some $0 \le a, b \le c$.*[3]

Proof. We begin by showing that all the nonabelian subgroups in the upper half of $\mathcal{L}(P)$ are isomorphic to some H_{ab}.

By Theorem 3.11, the maximal subgroups of P are isomorphic to H_{10} and H_{01}. Thereafter, as the order of the subgroups decreases, they are split and of the form $H_{ab} = P(m-a, n-b, c-a-b, 0)$ as long as $m-a \le n-b$. Suppose $m+k = n$ for some $k \ge 0$. The first possibility for a nonsplit subgroup to appear is when $m-a > n-b$, or $b-a > k$. Thus $H_{0,k+1}$ is the first subgroup of P that might have a nonsplit subgroup (none of the other H_{ab} of the same order satisfy $b-a = k+1$). However, $k+1 \ge c$ implies $H_{0,k+1}$ is abelian so we have reached the lower half of the lattice and all subgroups in the upper half are split and isomorphic to some H_{ab}.

Next we must show that centralizers of all nonabelian subgroups of P are isomorphic to some $H_{a'b'}$ with $a'+b' > c$. Proposition 4.4 shows that the centralizers of those nonabelian subgroups of P exactly equal to H_{ab}, where $a+b < c$, are equal to $H_{c-b,c-a}$.

[3]These are not the only conditions on the metacyclic parameters under which the result of the theorem holds, but they are the most succinct.

Now suppose that $M \leq P$ is isomorphic to H_{ab}, where $a + b < c$, and let $M = P(m-a, n-b, c-a-b, 0) = \langle s, t\rangle$ as in (3-1). By Lemma 4.1, $Z(M) = \langle s^{p^{c-a-b}}, t^{p^{c-a-b}}\rangle$, which has order $p^{m+n-2c+a+b}$, and, therefore, must be equal to $C_P(M)$. Thus $C_P(M) = P(c-b,\ c-a,\ 0,\ 0) \cong H_{c-b,c-a}$. □

Example 4.8. In Figure 4 we illustrate the CD-lattice for $P(4, 6, 3, 0)$, where $m < n$ and $c \leq n - m + 1$. The weights on the edges indicate the number of isomorphic copies of a particular subgroup coming from a parent H_{ab} (for example, each copy of H_{10} has p subgroups isomorphic to H_{20}), and the numbers in parentheses indicate the number of distinct subgroups of a particular form (for example, even though there are $2p^2$ paths from P to H_{21}, there are only p^2 distinct copies of H_{21} in P). Except for the weights, one can see that $\mathcal{L}(P)$ looks like $\mathcal{BEK}(P)$.

5. Two complete CD-lattices

We begin by applying the theory of Section 3 to a particular family of split metacyclic p-groups whose Chermak–Delgado lattices we can determine completely.

Theorem 5.1. *The Chermak–Delgado lattice of $P(m, n, 1, 0)$ is as illustrated in Figure 5.*

Proof. Since $|Z(P)| = p^{m+n-2}$, we know the only subgroups in $\mathcal{L}(P)$ other than $Z(P)$ and P are the maximal proper subgroups, which coincide with the maximal abelian subgroups on the line of symmetry at height p^{m+n-1}. From Theorem 3.11 we have

(1) $L = \langle x^p, y\rangle = P(m-1, n, 0, 0) \cong \mathbb{Z}_{p^{m-1}} \times \mathbb{Z}_p$,

(2) $M_i = \langle x^p, x^i y\rangle = \begin{cases} P(m-1, n, 0, 0) \cong \mathbb{Z}_{p^{m-1}} \times \mathbb{Z}_{p^n} & \text{if } m \leq n, \\ P(m, n-1, 0, 0) \cong \mathbb{Z}_{p^m} \times \mathbb{Z}_{p^{n-1}} & \text{if } m > n, \end{cases}$

(3) $R = \langle x, y^p\rangle = P(m, n-1, 0, 0) \cong \mathbb{Z}_{p^m} \times \mathbb{Z}_{p^{n-1}}$.

Note that if $n = 1$, then $M_i = \langle x^i y\rangle \cong \mathbb{Z}_{p^m}$ and $R = \langle x\rangle \cong \mathbb{Z}_{p^m}$. □

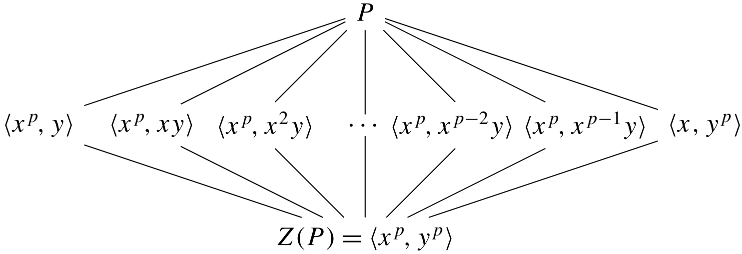


Figure 5. The Chermak–Delgado lattice of $P(m, n, 1, 0)$.

The theorem above shows that the CD-lattice of $P(m,n,1,0)$ is a *quasiantichain* in the sense of [Brewster et al. 2014b]. Indeed, while $\mathcal{L}(P)$ contains many intervals that are quasiantichains, the work in Section 3 shows that the whole CD-lattice of $P(m,n,c,0)$ is a quasiantichain (of width $p+1$) if and only if $c=1$.

The next example shows how some nonsplit subgroups can appear in $\mathcal{CD}(P)$, but modulo these irregular groups the CD-lattice looks like the BEK-lattice.

We build the CD-lattice of $P = P(6,5,4,0)$ one level at a time, using Theorems 3.11 and 3.12. Lowercase letters such as m, n, c, and s will always refer to the original group P, while uppercase letters such as M, N, C, and S will refer to the parameters of the particular subgroup in question:

- *Order* p^{10}. Since $m > n > m-c+1$, there are $p-1$ subgroups of type $J_1 = P(5,5,3,1)$ together with H_{10} and H_{01} at the maximal level.

- *Order* p^9. Since $H_{10} = P(5,5,3,0)$ has $M \le N$, it has p subgroups isomorphic to type L and one of type R. That is, we have p copies of H_{20} and one of H_{11}.

  Since $H_{01} = P(6,4,3,0)$ has $M > N$ and $N \le M-C+1$, it has p subgroups isomorphic to type R and one of type L. That is, we have p copies of H_{02} and one of H_{11}.

  Since $J_1 = P(5,5,3,1)$ has $S < C-1$, it has p subgroups isomorphic to type $L = P(4,5,2,1) = J_2$. Since $S = M-N+1$, we know J_1 has one subgroup of type $R = P(5,4,2,0) = H_{11}$.

  There are only p^2+p+1 subgroups of P of order p^9, so there must be intersections among the p^2+2p+1 subgroups listed above. The intersections can be hard to track because we often use [King 1973] to show the existence of alternative generators having nice properties. Although King shows how to construct the alternative generators, tracking all of them is a mind-numbing task that we will not illustrate.

- *Order* p^8. Since $H_{20} = P(4,5,2,0)$ has $M \le N$, it has p subgroups isomorphic to type $L = P(3,5,1,0) = H_{30}$ and one of type $R = P(4,4,1,0) = H_{21}$.

  Since $H_{11} = P(5,4,2,0)$ has $M > N$ and $N \le M-C+1$, it has p subgroups isomorphic to type $R = P(5,3,1,0) = H_{12}$ and one of type $L = P(4,4,1,0) = H_{21}$.

  Since $H_{02} = P(6,3,2,0)$ has $M > N$ and $N \le M-C+1$, it has p subgroups isomorphic to type $R = P(6,2,1,0) = H_{03}$ and one of type $L = P(5,3,1,0) = H_{12}$.

  Since $J_2 = P(4,5,2,1)$ has $S = C-1$, it has p subgroups isomorphic to type $L = P(6,2,1,0) = H_{03}$. Since $M-N+1 < S = C-1$, we know J_2 has one subgroup of type $R = P(5,3,1,0) = H_{12}$.

- *Order* p^7. Since $H_{30} = P(3,5,1,0)$ has $M \le N$, it has p subgroups isomorphic to type $L = P(2,5,0,0) = H_{40}$ and one of type $R = P(3,4,0,0) = H_{31}$.

  Since $H_{21} = P(4,4,1,0)$ has $M \le N$, it has p subgroups isomorphic to type $L = P(3,4,0,0) = H_{31}$ and one of type $R = P(4,3,0,0) = H_{22}$.

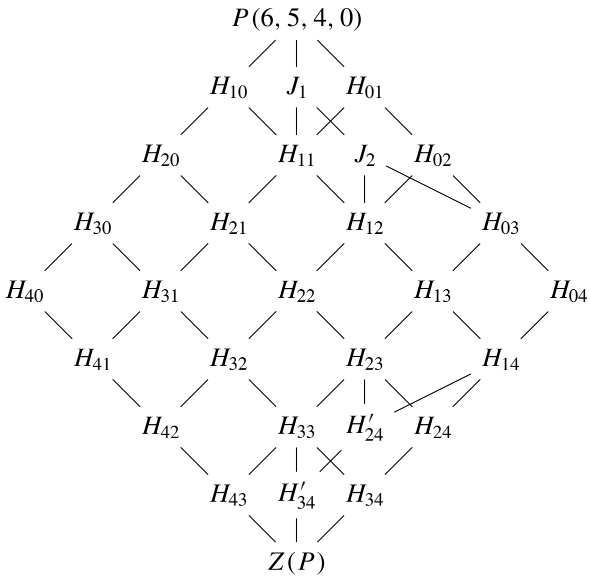


Figure 6. The collapsed CD-lattice of $P(6, 5, 4, 0)$.

Since $H_{12} = P(5, 3, 1, 0)$ has $M > N$ and $N \leq M - C + 1$, it has p subgroups isomorphic to type $R = P(5, 2, 0, 0) = H_{13}$ and one of type $L = P(4, 3, 0, 0) = H_{22}$.

Since $H_{03} = P(6, 2, 1, 0)$ has $M > N$ and $N \leq M - C + 1$, it has p subgroups isomorphic to type $R = P(6, 1, 0, 0) = H_{04}$ and one of type $L = P(5, 2, 0, 0) = H_{13}$.

• *Order* $\leq p^6$. Notice that all of the subgroups of order p^7 are abelian, so that is the row of maximal abelian subgroups of P. The rest of the CD-lattice consists of centralizers of the groups above. From Proposition 4.4 we know that $C_P(H_{ab}) = H_{c-b,c-a}$. Suppose $J_1 = P(5, 5, 3, 1) = \langle s, t\rangle$. Then Lemma 4.1 says that $Z(J_1) = \langle s^{p^3}, t^{p^3}\rangle = P(2, 2, 0, 1)$. This has order p^4, which is the order of $C_P(J_1)$, so $Z(J_1) = C_P(J_1)$. By the decision tree in Figure 2, we see that $Z(J_1)$ has an alternative presentation of the form $P(3, 1, 0, 0)$, making it isomorphic to H_{34}.

Similarly, the center of J_2 coincides with its centralizer in P and is of the form $P(2, 3, 0, 1)$. This subgroup has an alternative presentation as $P(4, 1, 0, 0)$, which is isomorphic to H_{24}.

A collapsed version of $\mathcal{L}(P)$ is shown in Figure 6, where multiple copies of subgroups are not indicated.

Acknowledgements

The authors would like to thank one of the reviewers for helpful comments that inspired us to delve deeper and write a better, more complete paper.

References

[An et al. 2015] L. An, J. P. Brennan, H. Qu, and E. Wilcox, "Chermak–Delgado lattice extension theorems", *Comm. Algebra* **43**:5 (2015), 2201–2213. MR Zbl

[Berkovich 2013] Y. Berkovich, "Maximal abelian and minimal nonabelian subgroups of some finite two-generator p-groups especially metacyclic", *Israel J. Math.* **194**:2 (2013), 831–869. MR Zbl

[Bidwell and Curran 2010] J. N. S. Bidwell and M. J. Curran, "Automorphisms of finite abelian groups", *Math. Proc. R. Ir. Acad.* **110A**:1 (2010), 57–71. MR Zbl

[Brewster and Wilcox 2012] B. Brewster and E. Wilcox, "Some groups with computable Chermak–Delgado lattices", *Bull. Aust. Math. Soc.* **86**:1 (2012), 29–40. MR Zbl

[Brewster et al. 2014a] B. Brewster, P. Hauck, and E. Wilcox, "Groups whose Chermak–Delgado lattice is a chain", *J. Group Theory* **17**:2 (2014), 253–265. MR Zbl

[Brewster et al. 2014b] B. Brewster, P. Hauck, and E. Wilcox, "Quasi-antichain Chermak–Delgado lattices of finite groups", *Arch. Math.* (*Basel*) **103**:4 (2014), 301–311. MR Zbl

[Chermak and Delgado 1989] A. Chermak and A. Delgado, "A measuring argument for finite groups", *Proc. Amer. Math. Soc.* **107**:4 (1989), 907–914. MR Zbl

[Davitt 1970] R. M. Davitt, "The automorphism group of a finite metacyclic p-group", *Proc. Amer. Math. Soc.* **25** (1970), 876–879. MR Zbl

[Dietz 1993] J. Dietz, "Stable splittings of classifying spaces of metacyclic p-groups, p odd", *J. Pure Appl. Algebra* **90**:2 (1993), 115–136. MR Zbl

[Glauberman 2006] G. Glauberman, "Centrally large subgroups of finite p-groups", *J. Algebra* **300**:2 (2006), 480–508. MR Zbl

[Héthelyi and Külshammer 2011] L. Héthelyi and B. Külshammer, "Characters, conjugacy classes and centrally large subgroups of p-groups of small rank", *J. Algebra* **340** (2011), 199–210. MR Zbl

[Isaacs 2008] I. M. Isaacs, *Finite group theory*, Graduate Studies in Mathematics **92**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[King 1973] B. W. King, "Presentations of metacyclic groups", *Bull. Austral. Math. Soc.* **8** (1973), 103–131. MR Zbl

[Mann 2010] A. Mann, "The number of subgroups of metacyclic groups", pp. 93–95 in *Character theory of finite groups*, Contemp. Math. **524**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Schulte 2001] M. Schulte, "Automorphisms of metacyclic p-groups with cyclic maximal subgroups", *Rose-Hulman Undergrad. Math J.* **2**:2 (2001).

Received: 2015-03-16 Revised: 2015-10-01 Accepted: 2015-10-27

erinbrush621@gmail.com St. Olaf College, Northfield, MN 55057, United States

dietz@stolaf.edu Department of Mathematics, Statistics and Computer Science, St Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057, United States

kjohnsontesch@gmail.com St. Olaf College, Northfield, MN 55057, United States

briannemp@gmail.com St. Olaf College, Northfield, MN 55057, United States

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.783

The left greedy Lie algebra basis and star graphs

Benjamin Walter and Aminreza Shiri

(Communicated by Vadim Ponomarenko)

We construct a basis for free Lie algebras via a *left greedy* bracketing algorithm on Lyndon–Shirshov words. We use a new tool — the configuration pairing between Lie brackets and graphs of Sinha and Walter — to show that the left greedy brackets form a basis. Our constructions further equip the left greedy brackets with a dual monomial Lie coalgebra basis of *star* graphs. We end with a brief example using the dual basis of star graphs in a Lie algebra computation.

1. Introduction

Lie algebras are classical objects with applications in differential geometry, theoretical physics, and computer science. A Lie algebra is a vector space which has an extra nonassociative (bilinear) operation called a *Lie bracket*, written $[a, b]$. The Lie bracket operation satisfies anticommutativity and Jacobi relations:

$$\text{(anticommutativity)} \qquad 0 = [a, b] + [b, a],$$
$$\text{(Jacobi)} \qquad 0 = [a, [b, c]] + [c, [a, b]] + [b, [c, a]].$$

A free Lie algebra is a Lie algebra whose bracket operation satisfies no extra relations — only the two written above and any relations which can be generated by combining them together. For example,

$$[a, [b, c]] - [[a, b], c] = [[c, a], b]$$

is a relation for free Lie algebras since $[c, [a, b]] = -[[a, b], c]$ by anticommutativity, and similarly for $[b, [c, a]]$. Free Lie algebras are fundamental in that every Lie algebra can be written *via generators and relations* as a free Lie algebra with extra relations placed on its bracket operation.

Recall that a set of elements generates an algebra if all other elements in the algebra can be obtained via sums of products of elements from the set. A minimal

MSC2010: primary 17B35; secondary 17B62, 16T15, 18D50.
Keywords: Lie algebra, bases, free Lie algebra.
This work is the result of an undergraduate research project of Shiri at the Middle East Technical University, Northern Cyprus Campus during the summer and fall of 2014.

generating set is called an algebra basis. We are interested in linear bases for algebras — these are minimal sets consisting of an algebra basis along with enough products of these so that all further algebra elements can be reached using only sums.

The current work describes a new linear basis for free Lie algebras, along with a new method for finding, computing with, and proving theorems about general free Lie algebra bases. Our method uses the graph/tree pairing developed in [Sinha 2005; 2006], which yields a new way to describe Lie coalgebras via graphs as applied in [Sinha and Walter 2011] and explained further in [Walter 2010]. Since we introduce a new and different way to perform calculations in free Lie algebras, we give many detailed examples throughout. For readers interested in a history of bases of free Lie algebras, we suggest [Bokut and Chibrikov 2006, §3.2].

2. Notation and classical constructions

In this paper we will say *alphabet* for a collection of abstract letters (or variables). A *word* in an alphabet is a (noncommutative, associative) string (or product) of letters from the alphabet. An ordering on an alphabet (such as the standard alphabetical ordering in English) induces an ordering on words called the *lexicographical* (or dictionary) ordering. The cyclic permutations of a word are given by removing letters from the beginning of the word and moving them to the end.

Example 2.1. The cyclic permutations of the word $abcd$ are $bcda$, $cdab$, $dabc$. The cyclic permutations of the word $aaabb$ are $aabba$, $abbaa$, $bbaaa$ and $baaab$.

A word that is lexicographically strictly less than all its cyclic permutations is called a *Lyndon–Shirshov word* (or just a *Lyndon word*). For instance, $aaabb$ is a Lyndon–Shirshov word, but $aabba$ and $abbaa$ are not. Neither is $abab$, since it equals one of its cyclic permutations.

A linear basis for the free Lie algebra on an alphabet can be built on Lyndon–Shirshov words using *standard bracketing* [Reutenauer 1993]i (see [Melançon and Reutenauer 1989; Chibrikov 2006; Stöhr 2008] for other methods). Given a Lyndon–Shirshov word w, its standard bracketing $[w]$ is recursively defined by $[w] = w$ if w has length 1, and by $[w] = [uv] = [[u], [v]]$ if v is a maximally long Lyndon–Shirshov suffix of $w = uv$, with u nonempty.

Example 2.2. The standard bracketing of $aaabb$ is

$$\begin{aligned}[aaabb] &= \big[[a], [aabb]\big] = \big[a, [\,[a], [abb]]\big] \\ &= \big[a, [a, [[ab], [b]]]\big] = \big[a, [a, [[a, b], b]]\big].\end{aligned}$$

The collection of all standard bracketings of Lyndon–Shirshov words gives a linear basis for the free Lie algebra on the underlying alphabet.

The standard bracketing of a Lyndon–Shirshov word can also be described recursively from the innermost brackets to the outermost roughly as follows: Read the letters of a Lyndon–Shirshov word from right to left looking for the first occurrence of consecutive letters, $\ldots a_i a_{i+1} \ldots$, where $a_i < a_{i+1}$ (called the *rightmost inversion*). Replace $a_i a_{i+1}$ by the bracket $[a_i, a_{i+1}]$, which we will consider to be a new *letter* placed in the ordered alphabet in the lexicographical position of the word $a_i a_{i+1}$. Repeat. (For a more detailed description see [Melançon and Reutenauer 1989, §2].)

Our construction of left greedy brackets will also proceed from the innermost bracket the to outermost bracket, similar to the *rewriting system* presented above. However, just as with the standard bracketing, left greedy brackets can also be described from the outermost to the innermost brackets.

3. Left greedy brackets and star graphs

Simple words.

Definition 3.1. Given a fixed letter a in an alphabet, an a-simple word is a word of the form $w = aa \cdots ax$ (written $w = a^n x$ for short), where x is any single letter not equal to a. The single-letter word $w = x$ (i.e., $w = a^0 x$) is also an a-simple word (for $x \neq a$).

The collection of all words in an alphabet is itself an (infinite) ordered alphabet (with the lexicographical ordering). A word in the alphabet whose letters are words in another alphabet will be casually referred to as a *word of words*. Note that such an expression of a word as a product of subwords is equivalent to a partition of the word. Considering words as ordered sets of letters, partitions are order-preserving surjections of sets; hence our notation for partitioning a word will be a double-headed arrow, $\twoheadrightarrow$, as defined below.

Remark 3.2. A partition of a word is equivalent to a rewriting which combines multiple subwords in parallel (compare [Mélançon 1997]). For our construction and proofs we will critically make use of the levels of nesting of partitions. We use the term *partition* so that our notation and our terminology reflect this emphasis.

Definition 3.3. A simple partition of a word w is an expression of w as subwords $w = \alpha_1 \alpha_2 \cdots \alpha_k$ where each α_i is an a-simple word and a is the first letter of w. We will write $w \twoheadrightarrow \alpha_1 \alpha_2 \cdots \alpha_k$.

Note that words have at most one simple partition. The subword α_1 must consist of the initial string of a's as well as the first letter of w other than a. If the letter in w following α_1 is a, then α_2 must consist of the next string of a's as well as the next letter other than a. If the letter following α_1 is not a, then α_2 will consist of only that one letter. (See the first line of Example 3.6.)

Remark 3.4. A simple partition of a word is equivalent to performing Lazard elimination [Lothaire 1997, Chapter 5] on the word, eliminating the first letter a via the bisection $(a^*(A\backslash a), a)$. It appears likely that the constructions of nested partitions and fully partitioned words which will follow may also be performed via a recursive series of Lazard elimination steps along the lines of: Order all words lexicographically. Eliminate them, one at time, beginning with the least ordered word a. (Possibly it will be best to restrict to words of length $\leq n$ at first.)

This would give an alternate proof that the left greedy brackets form a basis. However, making the previous statement precise and showing that it gives a well-defined recursive operation which will terminate is complicated. Also, following such a path would not yield the dual basis of star graphs, which we wish to exploit in later work.

Given a simple partition $w \twoheadrightarrow \alpha_1\alpha_2\cdots\alpha_k$, we may recurse: The a-simple subwords α_1, α_2, etc. are letters in the alphabet of words. They may have a further simple partition (now as α_1-simple words). This process constructs a unique nested partition of a word such that each nested level is a simple partition.

Definition 3.5. A word *fully partitions* if it has a series of simple partitions,

$$w \twoheadrightarrow \omega_1 \twoheadrightarrow \cdots \twoheadrightarrow \omega_\ell,$$

where ω_ℓ is the trivial coarse partition.

Colloquially, a word fully partitions if it is a simple word of simple words of simple words etc.

Example 3.6. Words fully partition as follows (for clarity we will use distinct delimiters (, [and { to indicate different nested levels of partition):

- $aaaab \twoheadrightarrow (aaaab)$.
- $ababb \twoheadrightarrow (ab)\,(ab)\,(b) \twoheadrightarrow [(ab)(ab)(b)]$.
- $aabcb \twoheadrightarrow (aab)\,(c)\,(b) \twoheadrightarrow [(aab)(c)]\,[(b)] \twoheadrightarrow \big\{[(aab)(c)]\,[(b)]\big\}$.
- $ababbabaab \twoheadrightarrow (ab)\,(ab)\,(b)\,(ab)\,(aab) \twoheadrightarrow [(ab)(ab)(b)]\,[(ab)(aab)]$
  $\twoheadrightarrow \big\{[(aab)(b)]\,[(ab)(aab)]\big\}$.

For visual clarity, we have found that indicating nested partitions via underlining is often more understandable than using nested parentheses:

$$aaaab \twoheadrightarrow \underline{aaaab}, \qquad ababb \twoheadrightarrow \underline{\underline{ab}\;\underline{ab}\;\underline{b}},$$

$$aabcb \twoheadrightarrow \underline{\underline{\underline{aab}\;\underline{c}}\;\underline{\underline{b}}}, \qquad ababbabaab \twoheadrightarrow \underline{\underline{\underline{ab}\;\underline{ab}\;\underline{b}}\;\underline{\underline{ab}\;\underline{aab}}},$$

$$abcabcabbabcaab \twoheadrightarrow \underline{\underline{\underline{\underline{ab}\;\underline{c}}\;\underline{\underline{ab}\;\underline{c}}\;\underline{\underline{ab}\;\underline{b}}}\;\underline{\underline{\underline{ab}\;\underline{c}}\;\underline{\underline{aab}}}}.$$

Example 3.7. Some words do not fully partition:

- $aaaa$ contains repetitions of only one letter.
- $aaba$ has the same initial and final letter.
- $abab \twoheadrightarrow (ab)(ab)$, which is a repetition of a single subword (ab).
- $abaabab \twoheadrightarrow (ab)(aab)(ab)$, which has the same initial and final subword (ab).
- $ababbcababb \twoheadrightarrow [(ab)(ab)(b)][(c)][(ab)(ab)(b)]$, which has the same initial and final subword $[(ab)(ab)(b)]$.

The following simple lemma follows immediately from standard facts about Lyndon–Shirshov words. We give a proof below for completeness.

Lemma 3.8. *Every Lyndon–Shirshov word fully partitions.*

Proof. Fix a word w with initial letter a. The only obstacle to the partition of w into a-simple words is whether the final letter and the initial letter match. More generally, each step of the recursive partitioning can be completed as long as the initial and final subword do not match. This fails only if the word has the form $w = \alpha\chi\alpha$, where α and χ are subwords (the subword χ may be empty and is likely not simple).

However no Lyndon–Shirshov word has this form. If χ is empty then $w = \alpha\alpha$, which is not Lyndon–Shirshov. If χ is nonempty, then one of the cyclic reorderings of w is lexicographically lower: either $\alpha\alpha\chi < \alpha\chi\alpha$ (if $\alpha < \chi$) or else $\chi\alpha\alpha < \alpha\chi\alpha$ (if $\chi < \alpha$). $\square$

Remark 3.9. Many non-Lyndon–Shirshov words also fully partition. The requirement that $w \neq \alpha\chi\alpha$ for any subwords α and χ is much weaker than the Lyndon–Shirshov requirement. Via some experimentation, we have found that it is possible to use methods similar to those presented in the current work to find new bases for Lie algebras which are constructed from sets of words other than the Lyndon–Shirshov words. It is unclear if these sets of words are also bases for the shuffle algebra.

Left greedy brackets.

Definition 3.10. The left greedy bracketing of the a-simple word $w = a^n x$, which we denote by $\llbracket w \rrbracket$, is $\llbracket a^n x \rrbracket = [a, [a, \cdots [a, [a, x]] \cdots]]$, the standard right-normed bracketing. The left greedy bracketing of a simple word of simple words (and, more generally, any fully partitioned word) is defined recursively:

$$\llbracket \alpha^n \chi \rrbracket = \big[\llbracket \alpha \rrbracket, [\llbracket \alpha \rrbracket, \dots [\llbracket \alpha \rrbracket, [\llbracket \alpha \rrbracket, \llbracket \chi \rrbracket]] \cdots]\big].$$

Example 3.11. Following are some examples of left greedy bracketings of fully partitioned words. To aid understanding in the examples below, we underline to indicate their full partition into simple words. (Note that we do not require words to be Lyndon–Shirshov in order to define their left greedy bracketing.)

- $\llbracket \underline{aaab} \rrbracket = \big[a,\ [a,\ [a,\ b]]\big]$.
- $\llbracket\, \underline{\underline{ab}\,\underline{ab}\,\underline{b}}\, \rrbracket = \big[[a, b], [[a, b],\ b]\big]$.
- $\llbracket\, \underline{\underline{\underline{aab}\,\underline{c}}\,\underline{b}}\, \rrbracket = \big[[[a, [a, b]], c], b\big]$.
- $\llbracket\, \underline{\underline{\underline{ab}\,\underline{ab}\,\underline{b}}\,\underline{\underline{ab}\,\underline{aab}}}\, \rrbracket = \big[[[a, b], [[a, b],\ b]],\ [[a, b],\ [a, [a, b]]]\big]$.

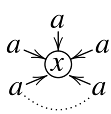
Remark 3.12. The name *left greedy* is due to the fact that the bracketing of the word $aaabcd$ begins with innermost bracket $[a, b]$ and then brackets leftwards i.e., $[a, [a, [a, b]]]$ before bracketing to the right. An alternative *right greedy* bracketing, would go to the right i.e., $[[[a, b], c], d]$ before bracketing leftwards. Both of these yield free Lie algebra bases, but the left greedy bracketing has a cleaner basis proof and appears to have better properties. We leave the discussion of the beneficial properties of the left greedy bracketing to a later paper.

Remark 3.13. Left greedy bracketing of Lyndon–Shirshov words is different than other bracketing methods considered in the literature. We give a few examples for comparison with other methods. Consider the Lyndon–Shirshov word $w = aababb$:

- $\llbracket aababb \rrbracket = \big[[[a, [a, b]],\ [a, b]],\ b\big]$, the left greedy bracketing.
- $[aababb] = \big[a, [[a, b],\ [[a, b],\ b]]\big]$, the standard Lyndon–Shirshov bracketing [Reutenauer 1993].
- $[aababb] = \big[[a, [a, b]],\ [[a, b],\ b]\big]$, the bracketing of [Chibrikov 2006, §4].

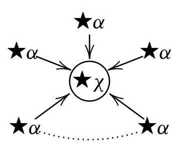
Star graphs. By a *graph* we mean a finite directed graph whose vertices are labeled by letters.

Definition 3.14. The star graph of the a-simple word $w = a^n x$, denoted $\star(w)$, is the graph with n vertices: labeled a, one vertex labeled x, and an edge from each a vertex to the vertex x:

$\star(a^n x) =$ 

.

The vertex x is called the *anchor vertex*. The star graph of a simple word of simple words (and, more generally, any fully partitioned word) is defined recursively. The graph $\star(\alpha^n \chi)$ consists of n disjoint subgraphs $\star(\alpha)$ and one disjoint subgraph $\star(\chi)$ with edges connecting the anchor vertices of the $\star(\alpha)$ to the anchor vertex of $\star(\chi)$:

$\star(\alpha^n \chi) =$ 

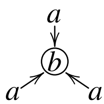
.

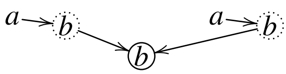
The anchor vertex of the subgraph $\star(\chi)$ serves as the anchor vertex of the star graph $\star(\alpha^n \chi)$.

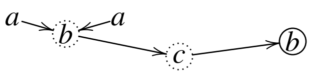
Remark 3.15. The star graph of an a-simple word consisting of one letter $w = x \neq a$ is a single anchor vertex:

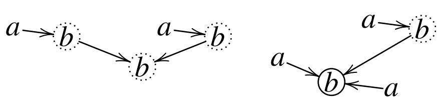
$$\bigstar(x) = \textcircled{x}.$$

Example 3.16. Following are some examples of star graphs. In the examples below, the anchor vertex of the subgraphs are indicated with dotted circles and the anchor vertex of the entire graph is indicated with a solid circle.

$\bigstar(\underline{aaab}) =$ 

$\bigstar(\underline{\underline{ab}\,\underline{ab}\,\underline{b}}) =$ 

$\bigstar(\underline{\underline{\underline{aab}\,\underline{c}}\,\underline{b}}) =$ 

$\bigstar(\underline{\underline{\underline{ab}\,\underline{ab}\,\underline{b}}\,\underline{\underline{ab}\,\underline{aab}}}) =$ 

Remark 3.17. The name *star graph* comes from imagining the graph $\bigstar(a^n b)$ as a sun (b) with planets (a) orbiting around it. The recursive construction of star graphs then composes suns and their planetary systems into orbiting star clusters, into galaxies, etc.

4. Configuration pairing

Throughout, assume that all graphs and Lie bracket expressions have labels and letters from the same alphabet.

Definition 4.1. Given a graph G and Lie bracket expression L, a bijection $\sigma : G \leftrightarrow L$ is a bijection between the vertices of G and the positions in L compatible with labels and letters (vertices of G are sent to positions in L labeled with the identical letter).

Example 4.2. Following are some basic examples investigating bijections between graphs and Lie bracket expressions:

- There are no bijections ${}_a\!\nearrow^{b}\!\searrow_a \leftrightarrow [a, b]$ because there are three vertices in the graph but only two positions in the Lie bracket expression. Similarly, there are no bijections ${}_a\!\nearrow^{b} \leftrightarrow [[a, b], a]$.
- There are no bijections ${}_a\!\nearrow^{b}\!\searrow_c \leftrightarrow [[a, b], a]$ because there is no letter c in the Lie bracket expression.

- There is only one bijection ${}_a\nearrow^b\searrow_c \leftrightarrow [[b,c],a]$ given by identifying each vertex with the correspondingly labeled position in the bracket expression.
- There are two bijections ${}_a\nearrow^b\searrow_a \leftrightarrow [[a,b],a]$ since there are two ways to choose an identification between the two vertices a and the two bracket positions a.
- More generally, there are $n!$ bijections $\bigstar(a^n b) \leftrightarrow [\![a^n b]\!]$.

Given a graph G and a subset V of the vertices of G, write $|V|$ for the full subgraph of G with vertices from V — i.e., two vertices are connected by an edge in $|V|$ if and only if they are connected by an edge in G. Recall that a graph is connected if every two vertices can be connected by a path of edges. We will say that directed graphs are connected if they are connected, ignoring edge directions.

The configuration pairing defined in [Sinha 2006] between directed graphs and rooted trees gives a pairing between graphs and Lie bracket expressions which can be defined as follows [Walter 2010].

Definition 4.3. Given a graph G and a Lie bracket expression L as well as a bijection $\sigma : G \leftrightarrow L$, the σ-configuration pairing of G and L is

$$\langle G, L\rangle_\sigma = \begin{cases} 0 & \text{if } L \text{ contains a subbracket expression } [H,K] \text{ so that the corresponding subgraphs } |\sigma^{-1}H| \text{ and } |\sigma^{-1}K| \text{ are not connected graphs with exactly one edge between them in } G, \\ (-1)^n & \text{otherwise (where } n \text{ is the number of edges of } G \text{ whose orientation corresponds under } \sigma \text{ to the right-to-left orientation of positions in } L). \end{cases}$$

The configuration pairing of G and L is the sum over all bijections σ,

$$\langle G, L\rangle = \sum_{\sigma : G \leftrightarrow L} \langle G, L\rangle_\sigma .$$

Casually, we will say that an edge ${}_a\nearrow^b$ in G whose orientation corresponds under σ to the right-to-left orientation of L (i.e., $\sigma(a)$ is to the right of $\sigma(b)$ in L) moves leftwards in L under σ.

Example 4.4. Following are some example computations of configuration pairings.

- $\big\langle {}_a\nearrow^b\searrow_c, [[b,c],a]\big\rangle = -1$. There is only one bijection. In this bijection only the edge ${}_a\nearrow^b$ moves leftwards in $[[b,c],a]$.
- $\big\langle {}_a\nearrow^a\searrow_b, [[a,b],a]\big\rangle = -1-1 = -2$.
- $\big\langle {}_a\nearrow^b\searrow_a, [[a,a],b]\big\rangle = 0$. For each of the two bijections, $\big|\sigma^{-1}\big([a,a]\big)\big|$ (the subgraph) is disconnected in G.

- $\big\langle {}_{a}\!\nearrow^{b}\!\searrow_{c}\!\nearrow^{a}, [[a,b],[a,c]]\big\rangle = -1+1 = 0$. There are two bijections. One bijection makes ${}_{c}\!\nearrow^{a}$ go leftwards. The other bijection makes ${}_{a}\!\nearrow^{b}$ and also ${}_{c}\!\nearrow^{a}$ go leftwards.
- The pairing of a linear (or *long*) graph ${}_{a_1}\!\nearrow^{a_2}\!\searrow_{\cdots}\!\nearrow^{a_n}$ with a bracket expression L is equal to the coefficient of the term $(a_1a_2\cdots a_n)$ in the associative polynomial for L [Walter 2010].

The configuration pairing encodes a duality between free Lie algebras and graphs modulo the Arnold and arrow-reversing identities [Sinha and Walter 2011]. In the current work we will use only that the configuration pairing is well-defined on Lie algebras — i.e., the configuration pairing vanishes on Jacobi and anticommutativity Lie bracket expressions. Thus the configuration pairing with graphs can be used to distinguish Lie bracket expressions, and in particular can be used to establish linear independence.

The main theorem will be proven essentially via recursive application of the following proposition, whose proof is trivial.

Proposition 4.5. *Let w_1 and w_2 be Lyndon–Shirshov words. If $w_1 = a^n b$ is a-simple then*

$$\big\langle \bigstar(w_1), \llbracket w_2 \rrbracket \big\rangle = \begin{cases} n! & \text{if } w_2 = w_1, \\ 0 & \text{otherwise.} \end{cases}$$

A similar result holds if w_2 is a-simple.

Proof. Suppose that w_1 and w_2 are Lyndon–Shirshov words with $\big\langle \bigstar(w_1), \llbracket w_2 \rrbracket \big\rangle \neq 0$. Note that w_1 and w_2 must be written with the same letters for any bijections $\sigma : \bigstar(w_1) \leftrightarrow \llbracket w_2 \rrbracket$ to exist. Furthermore w_1 and w_2 must share the same initial letter, since Lyndon–Shirshov words always begin with their lowest-ordered letter. Thus $w_1 = w_2$.

If $w_1 = w_2$, then there are $n!$ possible bijections $\sigma : \bigstar(a^n b) \leftrightarrow \llbracket a^n b \rrbracket$. For each of these $\big\langle \bigstar(a^n b), \llbracket a^n b \rrbracket \big\rangle_\sigma = 1$. □

5. The basis theorem

Theorem 5.1. *If w_1 and w_2 are Lyndon–Shirshov words, then $\big\langle \bigstar(w_1), \llbracket w_2 \rrbracket \big\rangle \neq 0$ if and only if $w_1 = w_2$ (in this case it is a product of factorials).*

Our desired result follows as a simple corollary.

Corollary 5.2. *The left greedy bracketing of Lyndon–Shirshov words gives a basis for free Lie algebras.*

Proof. A perfect pairing of graphs with left greedy brackets of Lyndon–Shirshov words implies that the left greedy brackets of Lyndon–Shirshov words are linearly independent. Since the number of Lyndon–Shirshov words of length n equals the dimension of the vector space of length-n Lie bracket expressions, this is enough to show that left greedy brackets of Lyndon–Shirshov words form a basis for the free Lie algebra. $\square$

Proof of Theorem 5.1. Suppose that w_1 and w_2 are Lyndon–Shirshov words with nonzero pairing:

$$\langle \bigstar(w_1), \llbracket w_2 \rrbracket \rangle \neq 0.$$

Fix a bijection $\sigma : \bigstar(w_1) \leftrightarrow \llbracket w_2 \rrbracket$. We will show that $w_1 = w_2$ by inducting on the depth of the nested partition resulting from fully partitioning the Lyndon–Shirshov words w_1 and w_2.

First note, as in the proof of Proposition 4.5, that w_1 and w_2 must be written with the same letters and must share the same initial letter, call it a. Thus w_1 and w_2 both fully partition, where the innermost partitions are a-simple words.

Write $w_2 \twoheadrightarrow (a^{n_1} b_1)(a^{n_2} b_2) \cdots (a^{n_k} b_k)$ for the innermost partition of w_2. According to its recursive definition, the bracket expression $\llbracket w_2 \rrbracket$ will have subbracket expressions $\llbracket a^{n_i} b_i \rrbracket$. From the definition of the configuration pairing, these must correspond under σ to connected, disjoint subgraphs of $\bigstar(w_1)$. However, the only possible connected subgraph of a star graph (with initial letter a) using the letters $a^{n_i} b_i$ is $\bigstar(a^{n_i} b_i)$. Note that this implies w_1 is composed of the subwords $(a^{n_i} b_i)$ (though possibly written in a different order). Furthermore, the first subword of w_1 must be $(a^{n_1} b_1)$ (just as in w_2), because Lyndon–Shirshov words must begin with their lexicographically least subword.

The induction step is equivalent to the previous case, treating subwords as letters. At the end of the previous case, for each simple subword u of w_2 the subbracket expressions $\llbracket u \rrbracket$ of $\llbracket w_2 \rrbracket$ correspond to disjoint connected subgraphs $\bigstar(u)$ of $\bigstar(w_1)$. Furthermore, the initial subword of w_2 coincides with the initial subword of w_1.

To finish the proof, we must note that $\langle \bigstar(w), \llbracket w \rrbracket \rangle \neq 0$ when w is a Lyndon–Shirshov word, since all bijections $\sigma : \bigstar(w) \leftrightarrow \llbracket w \rrbracket$ have $\langle \bigstar(w), \llbracket w \rrbracket \rangle_\sigma > 0$. In fact, a few short computations show that

$$\begin{aligned} \langle \bigstar(a^n b), \llbracket a^n b \rrbracket \rangle &= n!, \\ \langle \bigstar\big((a^{n_1} b_1)^m (a^{n_2} b_2)\big), \llbracket (a^{n_1} b_1)^m (a^{n_2} b_2) \rrbracket \rangle &= m!(n_1!)^m n_2!, \\ &\vdots \end{aligned}$$

$\square$

6. Projection onto the left greedy basis

Theorem 5.1 is of independent interest because it gives a direct, computational method for writing Lie bracket elements in terms of the left greedy Lyndon–Shirshov basis via projection.

Given a Lie bracket expression L, write $\{w_k\}$ for the set of Lyndon–Shirshov words written using the letters in L (with multiplicity). Left greedy brackets of Lyndon–Shirshov words form a linear basis, so it is possible to write L as a linear combination of the $\llbracket w_k \rrbracket$:

$$L = c_1 \llbracket w_1 \rrbracket + \cdots + c_n \llbracket w_n \rrbracket.$$

We may compute the constants c_k by pairing with $\star(w_k)$ since $\langle \star(w_k), \llbracket w_j \rrbracket \rangle = 0$ for $j \neq k$ by Theorem 5.1. This proves the following.

Corollary 6.1. *Given a Lie bracket expression L,*

$$L = \sum_{\substack{\text{Lyndon–Shirshov} \\ \text{words } w}} \frac{\langle \star(w), L \rangle}{\langle \star(w), \llbracket w \rrbracket \rangle} \llbracket w \rrbracket.$$

Recall that the denominators $\langle \star(w), \llbracket w \rrbracket \rangle$ are products of factorials. Interestingly, each coefficient in the expression above must be an integer (despite their large denominators).

Pairing computations are aided by the bracket/cobracket compatibility property of the configuration pairing. Bracket/cobracket compatibility states that pairings of a graph G with a bracket expression $[L, K]$ may be computed by calculating pairings of L and K with the subgraphs obtained by cutting G into two pieces by removing an edge. The following is Proposition 3.14 of [Sinha and Walter 2011].

Proposition 6.2. *Bracketing Lie expressions is dual to cutting graphs*:

$$\langle G, [H, K] \rangle = \sum_e \langle G_1^{\hat{e}}, H \rangle \cdot \langle G_2^{\hat{e}}, K \rangle \; - \; \langle G_1^{\hat{e}}, K \rangle \cdot \langle G_2^{\hat{e}}, H \rangle,$$

Where $G_1^{\hat{e}}$ and $G_2^{\hat{e}}$ are the graphs obtained by removing edge e from G, ordered so that e pointed from $G_1^{\hat{e}}$ to $G_2^{\hat{e}}$ in G.

Remark 6.3. Applying bracket/cobracket duality and the definition of the configuration pairing yields a recursive method for computation of $\langle G, L \rangle$. Consider the outermost bracketing $L = [H, K]$. Look for edges of G which can be removed to separate G into subgraphs $G_1^{\hat{e}}$ and $G_2^{\hat{e}}$ whose sizes matches that of H and K, and check that the subgraphs are written using the same letters as H and K. If this is not possible, then the bracketing is 0. Otherwise the bracketing is given by summing $\langle G_1^{\hat{e}}, H \rangle \cdot \langle G_2^{\hat{e}}, K \rangle$ (or the negative $-\langle G_1^{\hat{e}}, K \rangle \cdot \langle G_2^{\hat{e}}, H \rangle$ if e pointed so that $G_1^{\hat{e}}$ corresponded to K instead of H) over all such edges. Recurse. Note that removing an edge from a star graph will always result in subgraphs which are themselves star graphs (though possibly not star graphs of Lyndon–Shirshov words).

Example 6.4. Consider the Lie bracket expression $L = \big[[[a, b], b], [[a, b], a]\big]$. There are three Lyndon–Shirshov words with the letters $aaabbb$. These words, along with their partition, left greedy bracketings, and values of $\langle \star(w), \llbracket w \rrbracket \rangle$ are:

- $aaabbb$ which, partitions as $\underline{\underline{\underline{aaab}\,\underline{b}}\,\underline{b}}$ with

$$\llbracket aaabbb \rrbracket = \big[[[a,[a,[a,b]]],\ b],\ b\big] \quad \text{and} \quad \langle \star(aaabbb),\, \llbracket aaabbb \rrbracket \rangle = 3!.$$

- $aababb$ which, partitions as $\underline{\underline{\underline{aab}\,\underline{ab}}\,\underline{b}}$ with

$$\llbracket aababb \rrbracket = \big[[[a,[a,b]],\ [a,b]],\ b\big] \quad \text{and} \quad \langle \star(aababb),\, \llbracket aababb \rrbracket \rangle = 2!.$$

- $aabbab$ which, partitions as $\underline{\underline{\underline{aab}\,\underline{b}}\,\underline{ab}}$ with

$$\llbracket aabbab \rrbracket = \big[[[a,[a,b]],\ b],\ [a,b]\big] \quad \text{and} \quad \langle \star(aabbab),\, \llbracket aabbab \rrbracket \rangle = 2!.$$

The configuration pairings with L are as follows:

- $\big\langle \star(aaabbb),\, \big[[[a,b],b],[[a,b],a]\big]\big\rangle = 0$, because no edge of $\star(aaabbb)$ can be removed to separate it into subgraphs one of which has a single a and two b's (corresponding to the subbracket $\big[[a,b],b\big]$).
- $\big\langle \star(aababb),\, \big[[[a,b],b],[[a,b],a]\big]\big\rangle = 2$, because only the edge connecting $\star(aab)$ to the remainder of the graph cuts $\star(aababb)$ appropriately. This reduces the computation to

$$-\big\langle \star(aab),\, \big[[a,b],a\big]\big\rangle \cdot \big\langle \star(abb),\, \big[[a,b],b\big]\big\rangle = -(-2)\cdot 1 = 2.$$

- $\big\langle \star(aabbab),\, \big[[[a,b],b],[[a,b],a]\big]\big\rangle = -2$, because only the edge connecting $\star(aab)$ to the remainder of the graph cuts $\star(aabbab)$ appropriately. The computation reduces similarly.

Thus $L = \llbracket aababb \rrbracket - \llbracket aabbab \rrbracket$.

Acknowledgements

We would like to thank the reviewer for a careful and detailed reading of our submission and also for suggesting that we consider an application of Lazard elimination (see Remark 3.4) and rewriting systems on Lyndon words (see Remark 3.2).

References

[Bokut and Chibrikov 2006] L. Bokut and E. S. Chibrikov, "Lyndon-Shirshov words, Gröbner–Shirshov bases, and free Lie algebras", pp. 17–39 in *Non-associative algebra and its applications*, edited by L. Sabinin et al., Lect. Notes Pure Appl. Math. **246**, Chapman & Hall/CRC, Boca Raton, FL, 2006. MR 2203694 Zbl 1330.17001

[Chibrikov 2006] E. S. Chibrikov, "A right normed basis for free Lie algebras and Lyndon–Shirshov words", *J. Algebra* **302**:2 (2006), 593–612. MR 2293773 Zbl 1204.17006

[Lothaire 1997] M. Lothaire, *Combinatorics on words*, 2nd ed., Cambridge University Press, 1997. MR 1475463 Zbl 0874.20040

[Mélançon 1997] G. Mélançon, "Combinatorics of bases of free Lie algebras", 1997.

[Melançon and Reutenauer 1989] G. Melançon and C. Reutenauer, "Lyndon words, free algebras and shuffles", *Canad. J. Math.* **41**:4 (1989), 577–591. MR 1012617 Zbl 0694.17003

[Reutenauer 1993] C. Reutenauer, *Free Lie algebras*, London Mathematical Society Monographs. New Series **7**, The Clarendon Press, Oxford University Press, New York, 1993. MR 1231799 Zbl 0798.17001

[Sinha 2005] D. Sinha, "A pairing between graphs and trees", preprint, 2005. arXiv 0502547

[Sinha 2006] D. P. Sinha, "Operads and knot spaces", *J. Amer. Math. Soc.* **19**:2 (2006), 461–486. MR 2188133 Zbl 1112.57004

[Sinha and Walter 2011] D. Sinha and B. Walter, "Lie coalgebras and rational homotopy theory, I: Graph coalgebras", *Homology Homotopy Appl.* **13**:2 (2011), 263–292. MR 2861231 Zbl 1231.55009

[Stöhr 2008] R. Stöhr, "Bases, filtrations and module decompositions of free Lie algebras", *J. Pure Appl. Algebra* **212**:5 (2008), 1187–1206. MR 2387595 Zbl 1192.17001

[Walter 2010] B. Walter, "Lie algebra configuration pairing", preprint, 2010. arXiv 1010.4732

Received: 2015-04-29 Revised: 2015-10-23 Accepted: 2015-10-27

benjamin@metu.edu.tr
Department of Mathematics, Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, KKTC, 10 Mersin, Turkey

amin.shiri@metu.edu.tr
Department of Mathematics, Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, KKTC, 10 Mersin, Turkey

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.797

Note on superpatterns

Daniel Gray and Hua Wang

(Communicated by Joshua Cooper)

Given a set P of permutations, a P-superpattern is a permutation that contains every permutation in P as a pattern. The study of the minimum length of a superpattern has been of interest. For P being the set of all permutations of a given length, bounds on the minimum length have been improved over the years, and the minimum length is conjectured to be asymptotic with k^2/e^2. Similar questions have been considered for the set of layered permutations. We consider superpatterns with respect to packing colored permutations or multiple copies of permutations. Some simple but interesting observations will be presented. We also propose a few questions.

1. Introduction

Given a permutation π of length n, a pattern σ is said to be contained in π, or σ *occurs* in π, if a subsequence of π is order isomorphic to σ. For instance, the permutation $\pi = 51342$ contains two occurrences of the pattern $\sigma = 321$ as the subsequences 532 and 542. Much effort has been devoted to the study of occurrences of patterns in a permutation, most of which involves studying permutations which avoid a particular pattern, i.e., pattern avoidance.

As a symmetric problem to pattern avoidance, the concept of a superpattern concerns packing all patterns from a given set into a single permutation.

Definition. Let P be a set of permutations. A P-superpattern is a permutation that contains π for every $\pi \in P$.

Superpatterns were first introduced in [Arratia 1999]. The natural question immediately following this definition is to find the minimum length of a P-superpattern. When P is the set of all permutations of length k, this minimum length is denoted by $\mathrm{sp}(k)$ and has been vigorously studied. The trivial upper bound of k^2 was improved to $\frac{2}{3}k^2$ in [Eriksson et al. 2002], and was conjectured to be asymptotic with $\frac{1}{2}k^2$. Later, it was shown in [Miller 2009] that $\mathrm{sp}(k) \le \frac{1}{2}k(k+1)$ through

MSC2010: 05A05.
Keywords: superpatterns, colored permutations.
This work was partially supported by grants from the Simons Foundation (#245307).

the construction of a zigzag k-superword. More recently, bounds on the minimum lengths of superpatterns containing all layered permutations were considered in [Gray 2015]. Constructions similar to that in [Miller 2009] were also used to study superpatterns containing simple patterns of length k in [Gray $\geq$ 2016].

Definition. An m-colored permutation χ of length n is a permutation of length n in which each element is assigned one of m distinct colors.

For example, let $\chi = 3_1 2_1 5_1 1_1 4_2$ be a 2-colored permutation, where 3, 2, 5, and 1 have color 1, while 4 has color 2. Analogous to the case of noncolored patterns, the colored pattern $\phi = 2_1 1_1 3_2$ occurs in χ as the subsequences $3_1 1_1 4_2$, $2_1 1_1 4_2$, and $3_1 2_1 4_2$, but not $3_1 2_1 5_1$.

Colored permutations are of interest in the study of patterns and pattern avoidance [Savage and Wilf 2006]. Packing densities of colored permutations were also recently considered [Just and Wang 2016].

In this note we consider superpatterns of different sets of colored permutations. Some elementary, but interesting, observations will be presented. We also propose some questions from these studies.

2. Superpatterns containing all colored permutations

Let $S(k, m)$ denote the set of m-colored permutations of length k and

$$\mathrm{sp}(k, m) = \min\{|p| : p \text{ is an } S(k, m)\text{-superpattern}\}.$$

The following presents a simple connection between $\mathrm{sp}(k, m)$ and $\mathrm{sp}(k)$.

Theorem 2.1. *For any positive integers k and m, we have*

$$\mathrm{sp}(k, m) = m \cdot \mathrm{sp}(k).$$

Proof. Let p' be an $S(k, m)$-superpattern, and denote by p'_i the subsequence of p' of color i (for any $1 \leq i \leq m$). Then p'_i, without the color, is a superpattern containing all noncolored patterns of length k. Consequently $|p'_i| \geq \mathrm{sp}(k)$ for any i and

$$|p'| = \sum_{i=1}^{m} |p'_i| \geq m \cdot \mathrm{sp}(k).$$

On the other hand, let p be a permutation of length $\mathrm{sp}(k)$ that contains all noncolored patterns of length k. Construct an m-colored permutation p'' from p by replacing each $1 \leq j \leq \mathrm{sp}(k)$ in p by the sequence

$$s_j := [m(j-1)+1]_1 [m(j-1)+2]_2 \cdots [m(j-1)+m]_m.$$

Note that $|p''| = m \cdot |p| = m \cdot \mathrm{sp}(k)$. For any pattern $\pi \in S(k, m)$, the noncolored version is contained in p and the corresponding colored pattern can be found in p'' by choosing corresponding digits in s_j with the required color. Thus,

$$\mathrm{sp}(k, m) \leq |p''| = m \cdot \mathrm{sp}(k). \qquad \square$$

For example, $p = 132$ is a superpattern containing all patterns of length 2, and p is of length $\mathrm{sp}(2) = 3$. An $S(2, 3)$-superpattern p'' can be constructed as

$$1_1 2_2 3_3 7_1 8_2 9_3 4_1 5_2 6_3.$$

As an immediate consequence of Theorem 2.1, the established asymptotic bounds for $\mathrm{sp}(k)$ apply directly to $\mathrm{sp}(k, m)$. The trivial asymptotic lower bound k^2/e^2 for $\mathrm{sp}(k)$ follows from

$$\binom{\mathrm{sp}(k)}{k} \geq k!$$

and a standard application of Stirling's approximation for factorials [Arratia 1999].

Corollary 2.2. *For any positive integers k and m,*

$$mk^2/e^2 \leq \mathrm{sp}(k, m) \leq \tfrac{1}{2} mk(k+1).$$

Remark. The arguments in Theorem 2.1 establish the same relationship between the colored and noncolored versions of superpatterns containing any particular subset of the length-k permutations, such as the layered permutations [Gray 2015] and simple and alternating permutations [Gray $\geq$ 2016], and consequently provide bounds on the minimum lengths of these colored superpatterns.

3. Monochromatic and nonmonochromatic patterns

Let $NMS(k, m)$ be the set of nonmonochromatic m-colored patterns of length k and $MS(k, m)$ be the set of all monochromatic m-colored patterns of length k. Then, $S(k, m)$ is the disjoint union of $NMS(k, m)$ and $MS(k, m)$. It is easy to see that

$$|MS(k, m)| = mk!,$$

and consequently,

$$\begin{aligned} |NMS(k, m)| &= |S(k, m)| - |MS(k, m)| = m^k k! - mk! = (m^k - m)k! \\ &= (m^{k-1} - 1)\,|MS(k, m)|. \end{aligned}$$

Given any $NMS(k, m)$-superpattern of length n, we must have

$$\binom{n}{k} \geq (m^k - m)k!,$$

implying (by way of a standard application of Stirling's approximation for factorials)

$$n \geq mk^2/e^2,$$

the same asymptotic lower bound for $\mathrm{sp}(k, m)$ for general $S(k, m)$-superpatterns.

Letting

$$\mathrm{nmsp}(k, m) = \min\bigl\{|p| : p \text{ is an } NMS(k, m)\text{-superpattern}\bigr\},$$

we have the simple consequence that

$$mk^2/e^2 \leq \mathrm{nmsp}(k, m) \leq \mathrm{sp}(k, m) \leq \tfrac{1}{2} mk(k+1). \tag{1}$$

On the other hand, exactly the same argument as that of Theorem 2.1 implies

$$\operatorname{msp}(k, m) = m \cdot \operatorname{sp}(k), \tag{2}$$

where

$$\operatorname{msp}(k, m) = \min\big\{|p| : p \text{ is an } MS(k, m)\text{-superpattern}\big\}.$$

Remark. Equations (1) and (2) imply, in addition to the semitrivial bounds of $\operatorname{msp}(k, m)$ and $\operatorname{nmsp}(k, m)$, that

$$\operatorname{msp}(k, m) = m \cdot \operatorname{sp}(k) = \operatorname{sp}(k, m) \geq \operatorname{nmsp}(k, m), \tag{3}$$

a rather surprising fact given that $|NMS(k, m)| = (m^{k-1} - 1)\,|MS(k, m)|$.

While it may be a bit unexpected to see that $\operatorname{msp}(k, m) = \operatorname{sp}(k, m)$, a natural question follows.

Question 3.1. *Does strict inequality hold in (3)?*

In the special case for $k = 2$, the proposition below answers Question 3.1 in the affirmative.

Proposition 3.2. *For any positive integer m, we have*

$$3m = \operatorname{msp}(2, m) > 3m - 2 \geq \operatorname{nmsp}(2, m).$$

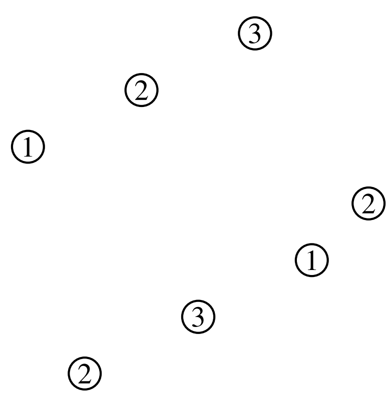
Proof. Clearly, $\operatorname{sp}(2) = 3$, and hence $\operatorname{msp}(2, m) = m \cdot \operatorname{sp}(2) = 3m$. Let p be a permutation of length $3m - 2$ defined as

$$[2m-1]_1 1_2 [2m]_2 2_3 \cdots [3m-3]_{m-1} [m-1]_m [3m-2]_m m_1 [m+1]_2 [m+2]_3 \cdots [2m-2]_{m-1}.$$

For instance, if $m = 3$, then

$$p = 5_1 1_2 6_2 2_3 7_3 3_1 4_2,$$

and the graph of p is depicted below:



In general, the graph of p will be the disjoint union of two increasing subsequences. The top row is of length m and the bottom row is of length $2m-2$, and every entry of the top row is larger than every entry of the bottom row. The entries of p alternate from the top row to the bottom row until there are m entries in the top row. Then, $m-2$ more entries are added to the bottom row. The i-th entry of the top row will have color i for $1 \le i \le m$, while the i-th entry of the bottom row will have color $i+1 \pmod{m}$.

For $i, j \in [1, m]$ with $i \neq j$, the pattern $1_i 2_j$ is contained in the bottom row, the only exception being the pattern $1_1 2_j$, which is contained in the top row. The pattern $2_i 1_j$ can be found by selecting the unique entry colored i from the top row, and taking an entry in the bottom row colored j which lies to the right of the entry just selected. Then, p is an $NMS(2, m)$-superpattern. □

To answer Question 3.1 in general appears to be very difficult. In an effort to further understand the relationship between monochromatic and nonmonochromatic superpatterns, we also point out the following.

Proposition 3.3. *For any positive integers $k \ge 2$ and m, we have*

$$\mathrm{msp}(k-1, m) \le \mathrm{nmsp}(k, m) \le \mathrm{msp}(k, m).$$

Proof. The second inequality is implied by the remark on page 800. To see the first inequality, let q be an m-colored pattern of length k whose first $k-1$ entries are colored by color i and whose k-th entry is colored by $j \neq i$. Then, the first $k-1$ entries of q form a monochromatic pattern of length $k-1$.

Since q is a nonmonochromatic m-colored pattern of length k, it must be contained in any $NMS(k, m)$-superpattern. Noting that we could have colored the first $k-1$ entries of q monochromatically using any of the m colors, any $NMS(k, m)$-superpattern must contain all monochromatic m-colored patterns of length $k-1$. Thus, $\mathrm{nmsp}(k, m) \ge \mathrm{msp}(k-1, m)$. □

4. Packing multiple copies of all patterns

The idea of superpatterns lies in the fact that they contain each permutation (from a given set of permutations) at least once. A natural generalization seems to be superpatterns that contain each permutation at least a given number of times.

Definition. For a given set P of permutations, a P_ℓ-superpattern is a permutation containing each pattern $\pi \in P$ at least ℓ times.

Define $\mathrm{sp}_\ell(k)$, $\mathrm{sp}_\ell(k, m)$, $\mathrm{msp}_\ell(k, m)$, and $\mathrm{nmsp}_\ell(k, m)$ accordingly. Some trivial facts follow immediately:

- $\mathrm{sp}_\ell(k, m) = m \cdot \mathrm{sp}_\ell(k)$. This can be seen by following exactly the same argument as that of Theorem 2.1.

- $\mathrm{sp}_\ell(k) \le \ell \cdot \mathrm{sp}(k)$ and $\mathrm{sp}_\ell(k,m) \le \ell \cdot \mathrm{sp}(k,m)$. For permutations p of length n and q of length m, the direct sum $p \oplus q$ is the permutation that has the first n entries from p and the next m entries from entries of q shifted by n. That is,

$$p \oplus q = p_1 p_2 \cdots p_n (q_1+n)(q_2+n)\cdots(q_m+n).$$

  Given a permutation p of length $\mathrm{sp}(k)$ that contains all patterns of length k, the permutation $\bigoplus_{i=1}^{\ell} p$ clearly contains each length-k pattern in each of the ℓ summands. Hence, $\mathrm{sp}_\ell(k) \le \ell \cdot \mathrm{sp}(k)$. A similar argument holds for $\mathrm{sp}_\ell(k,m) \le \ell \cdot \mathrm{sp}(k,m)$.

- $\mathrm{sp}_m(k) \le \mathrm{msp}(k,m)$. This can be seen from removing the colors of an $MS(k,m)$-superpattern.

The asymptotic lower bounds, for k large and ℓ constant, of $\mathrm{sp}_\ell(k)$ or $\mathrm{sp}_\ell(k,m)$ stay the same as $\mathrm{sp}(k)$ or $\mathrm{sp}(k,m)$. Given that the multiple copies of patterns need not be disjoint, it is natural to ask for improvement of the upper bounds above. The existing constructions (that provided upper bounds for the minimum lengths of various superpatterns) such as those in [Arratia 1999; Gray 2015; ≥ 2016] do not directly generalize to the case of packing multiple copies of every permutation. We conclude this note by showing a nontrivial upper bound for the sp_ℓ function.

Definition. For $k, n \in \mathbb{N}$, let $q = q_1q_2q_3\cdots q_k$ be a pattern of length k and let $w = w_1w_2w_3\cdots w_n$ be a word of length n. We say that q is "contained exactly in w" if there is a subsequence of length k, say

$$(w_{i_1}, w_{i_2}, \ldots, w_{i_k}),$$

such that $w_{i_j} = q_j$ for all $1 \le j \le k$.

Theorem 4.1. *For $k, \ell \in \mathbb{N}$, we have*

$$\mathrm{sp}_\ell(k) \le \begin{cases} \frac{1}{2}(k+1)(k+\ell-1) & \text{if } k \text{ is odd,} \\ \frac{1}{2}(k+1)(k+\ell-1) & \text{if } k \text{ is even and } \ell \text{ is odd,} \\ \frac{1}{2}(k+1)(k+\ell-1)+1 & \text{if } k \text{ is even and } \ell \text{ is even.} \end{cases}$$

Proof. We begin with Allison Miller's construction [2009] of the zigzag k-superword. Let k_o (resp. k_e) be the smallest odd (resp. even) integer at least as large as k. We make the following definitions:

$$\bar{k}_o = 1357\cdots k_o \quad \text{and} \quad \bar{k}_e = k_e \cdots 8642.$$

Define

$$w = \bar{k}_o\bar{k}_e\bar{k}_o\bar{k}_e\cdots\bar{k}_o\bar{k}_e$$

if k is even or

$$w = \bar{k}_o\bar{k}_e\bar{k}_o\bar{k}_e\cdots\bar{k}_o\bar{k}_e\bar{k}_o$$

if k is odd, where the combined number of copies of $\bar{k}_o$ and $\bar{k}_e$ is exactly k. The word w is what Miller calls the zigzag k-superword.

Let q be a pattern of length k, and let $q+1$ be the permutation of the set $\{2, 3, 4, \dots, k+1\}$ obtained by adding 1 to each entry of q. Number the runs of w from left to right in increasing order and let $m(q)$ be the number of runs needed (in $\bar{k}_o\bar{k}_e\bar{k}_o\bar{k}_e\cdots$) to contain q. Miller shows that

$$m(q)+m(q+1)\le 2k+1, \tag{4}$$

but in fact, the same steps can be used to show equality in (4). Hence, either $m(q)$ or $m(q+1)$ is at most k, which implies either q or $q+1$ is contained exactly in w.

Now, consider the finite word

$$w(\ell)=\begin{cases}\bar{k}_o\bar{k}_e\bar{k}_o\bar{k}_e\cdots\bar{k}_o\bar{k}_e & \text{if } k \text{ is even,}\\ \bar{k}_o\bar{k}_e\bar{k}_o\bar{k}_e\cdots\bar{k}_o\bar{k}_e\bar{k}_o & \text{if } k \text{ is odd,}\end{cases}$$

where the combined number of copies of $\bar{k}_o$ and $\bar{k}_e$ is exactly $k+\ell-1$. Suppose without loss of generality that $m(q)\le k$, and recall, $m(q+1)=2k+1-m(q)$. Since q is contained in the first $m(q)$ copies of $w(\ell)$, and each run of $w(\ell)$ is repeated every two times, there is another copy of q contained between the third run and the $(m(q)+2)$-th run, yet another copy of q contained between the fifth and $(m(q)+4)$-th runs, and so on for as long as we do not exceed $k+\ell-1$ runs. Thus, there are at least

$$1+\left\lfloor \tfrac{1}{2}\big((k+\ell-1)-m(q)\big)\right\rfloor=\left\lfloor \tfrac{1}{2}\big(\ell+(k-m(q))+1\big)\right\rfloor$$

copies of q contained exactly in $w(\ell)$. Hence, if $k-m(q)\ge \ell-1$, we successfully have at least ℓ copies of q. Then, let us suppose that $k-m(q)<\ell-1$. For the same reason as above, there are at least

$$1+\left\lfloor \tfrac{1}{2}\big((k+\ell-1)-(2k+1-m(q))\big)\right\rfloor=\left\lfloor \tfrac{1}{2}\big(\ell-(k-m(q))\big)\right\rfloor\ge 1$$

copies of $q+1$ contained exactly in $w(\ell)$. Hence, the combined number of copies of $m(q)$ and $m(q+1)$ is at least

$$\left\lfloor \tfrac{1}{2}\big(\ell+(k-m(q))+1\big)\right\rfloor+\left\lfloor \tfrac{1}{2}\big(\ell-(k-m(q))\big)\right\rfloor=\ell.$$

Finding a permutation p representing $w(\ell)$ is routine. Note that p will contain at least ℓ copies of q. Let us consider the length of $w(\ell)$. First suppose k is odd. Then, there are $\frac{1}{2}(k+1)$ entries each in $\bar{k}_o$ and $\bar{k}_e$. Miller shows that w is of length $\frac{1}{2}k(k+1)$, to which we add $\ell-1$ more runs. Hence, the length of $w(\ell)$ is

$$\tfrac{1}{2}k(k+1)+(\ell-1)\tfrac{1}{2}(k+1)=\tfrac{1}{2}(k+1)(k+\ell-1).$$

Now suppose that k is even. Then, there are $\frac{1}{2}k$ entries in $\bar{k}_e$ and $1+\frac{1}{2}k=\frac{1}{2}(k+2)$ entries in $\bar{k}_o$. If ℓ is odd, then we have added $\frac{1}{2}(\ell-1)$ copies each of $\bar{k}_e$ and $\bar{k}_o$.

Thus, the length of $w(\ell)$ is

$$\tfrac{1}{2}k(k+1)+\tfrac{1}{2}(\ell-1)\tfrac{1}{2}k+\tfrac{1}{2}(\ell-1)\tfrac{1}{2}(k+2)=\tfrac{1}{2}(k+1)(k+\ell-1).$$

If ℓ is even, then we have added $\frac{1}{2}(\ell-2)$ copies of $\bar{k}_e$ and $\frac{1}{2}\ell$ copies of $\bar{k}_o$. Therefore, the length of $w(\ell)$ is

$$\tfrac{1}{2}k(k+1)+\tfrac{1}{2}(\ell-2)\tfrac{1}{2}k+\tfrac{1}{2}\ell\tfrac{1}{2}(k+2)=\tfrac{1}{2}(k+1)(k+\ell-1)+1. \qquad \square$$

Remark. The above argument can be easily modified, by using the construction in Theorem 2.1, to provide less trivial upper bounds for $\mathrm{sp}_\ell(k,m)$.

Remark. It is also interesting to note that, if one takes a superpattern from $S(k,m)$ achieving $\mathrm{sp}(k,m)$ and removes colors, the resulting noncolored permutation is a superpattern that contains each k-pattern m^k times (since there are m^k different ways to color a k-pattern with m colors).

Acknowledgement

We greatly appreciate many helpful suggestions and corrections from the anonymous referee.

References

[Arratia 1999] R. Arratia, "On the Stanley–Wilf conjecture for the number of permutations avoiding a given pattern", *Electron. J. Combin.* **6** (1999), Note N1. MR 1710623 Zbl 0922.05002

[Eriksson et al. 2002] H. Eriksson, K. Eriksson, S. Linusson, and J. Wästlund, "Dense packing of patterns in a permutation", in *FPSAC '02: 14th Conference on Formal Power Series and Algebraic Combinatorics* (Melbourne 2002), 2002.

[Gray 2015] D. Gray, "Bounds on superpatterns containing all layered permutations", *Graphs Combin.* **31**:4 (2015), 941–952. MR 3357666 Zbl 1316.05001

[Gray $\geq$ 2016] D. Gray, "Simple permutations, alternating permutations, and the superpatterns that contain them", preprint. To appear in *Ars Combin.*

[Just and Wang 2016] M. Just and H. Wang, "Note on packing patterns in colored permutations", *Online J. Anal. Comb.* **11** (2016). Zbl 06574956

[Miller 2009] A. Miller, "Asymptotic bounds for permutations containing many different patterns", *J. Combin. Theory Ser. A* **116**:1 (2009), 92–108. MR 2469250 Zbl 1177.05012

[Savage and Wilf 2006] C. D. Savage and H. S. Wilf, "Pattern avoidance in compositions and multiset permutations", *Adv. in Appl. Math.* **36**:2 (2006), 194–201. MR 2199988 Zbl 1087.05002

Received: 2015-05-01 Revised: 2015-09-10 Accepted: 2015-09-17

dgray1@ufl.edu Department of Mathematics, University of Florida, Gainesville, FL 32611, United States

hwang@georgiasouthern.edu Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, United States

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.805

Lifting representations of finite reductive groups: a character relation

Jeffrey D. Adler, Michael Cassel,
Joshua M. Lansky, Emma Morgan and Yifei Zhao

(Communicated by Ken Ono)

Given a connected reductive group $\widetilde{G}$ over a finite field k, and a semisimple k-automorphism ε of $\widetilde{G}$ of finite order, let G denote the connected part of the group of ε-fixed points. Two of the authors have previously shown that there exists a natural lifting from series of representations of $G(k)$ to series for $\widetilde{G}(k)$. In the case of Deligne–Lusztig representations, we show that this lifting satisfies a character relation analogous to that of Shintani.

0. Introduction

Suppose k is a finite field, $\widetilde{G}$ is a connected reductive k-group, and ε is a semisimple k-automorphism of $\widetilde{G}$ of finite order ℓ. Let G be the connected part of the group $\widetilde{G}^{\varepsilon}$ of ε-fixed points of $\widetilde{G}$. We will see (Proposition 1) that G is also a connected reductive k-group. Two of the authors have constructed a natural lifting [Adler and Lansky 2014] from series of irreducible representations of $G(k)$ to analogous series for $\widetilde{G}(k)$. In certain cases, this lifting is known to coincide with the Shintani lifting (see [Gyoja 1979, Theorem 7.2], [Digne 1987, Corollary 3.6], or [Silberger and Zink 2005, Proposition B4.4], for example), so it is natural to ask whether there is a relation, analogous to that of Shintani, between the character of a representation π of $G(k)$ and the ε-twisted character of its lift $\tilde{\pi}$. The purpose of this note is to prove an affirmative answer in the case where π is a Deligne–Lusztig representation, irreducible or not.

Let $\widetilde{G}^*$ and G^* denote the duals of $\widetilde{G}$ and G. For each semisimple element $z \in G^*(k)$, one obtains a collection $\mathcal{E}_z(G(k))$ of irreducible representations of $G(k)$, and these collections, known as *rational Lusztig series*, partition the set $\mathcal{E}(G(k))$ of (equivalence classes of) irreducible representations of $G(k)$ [Lusztig 1984, §14.1]. For example, suppose that z is regular in G^*, and let $T^* \subseteq G^*$ be the unique maximal

MSC2010: primary 20C33; secondary 20G40.
Keywords: finite reductive groups, Deligne–Lusztig representations, liftings, character relations.

k-torus containing z. Then the pair (T^*, z) corresponds to a pair (T, θ), where $T \subseteq G$ is a maximal k-torus, and θ is a complex-valued character of $T(k)$. This latter pair is uniquely determined up to $G(k)$-conjugacy. The Lusztig series $\mathcal{E}_z(G(k))$ corresponding to z is the set of irreducible components of the Deligne–Lusztig virtual representation whose character is $\mathrm{R}_T^G\,\theta$. An earlier work [Adler and Lansky 2014, Corollary 11.3] presents a natural map from semisimple conjugacy classes in $G^*(k)$ to semisimple classes in $\widetilde{G}^*(k)$, thus lifting each Lusztig series for $G(k)$ to one for $\widetilde{G}(k)$. The series of representations coming from $\pm\mathrm{R}_T^G\theta$ lifts to that coming from $\pm\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\tilde{\theta}$, where we will see that $\widetilde{T} := C_{\widetilde{G}}(T)$ is a maximal k-torus in $\widetilde{G}$, and $\tilde{\theta} = \theta \circ \mathcal{N}$, where $\mathcal{N} : \widetilde{T} \to T$ is the norm map defined by

$$\mathcal{N}(t) = t\varepsilon(t) \cdots \varepsilon^{\ell-1}(t).$$

(Of course, one could define a similar map $\mathcal{N}$ on any ε-invariant torus in $\widetilde{G}$.) In order to understand better this lifting of representations, one would like to have a relation between the character $\mathrm{R}_T^G\,\theta$ (for θ an arbitrary character of $T(k)$, not necessarily associated to a regular element of $G^*(k)$) and the ε-twisted character $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_\varepsilon$ associated to the ε-invariant character $\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta}$. We prove that such a relation holds at sufficiently regular points, provided that one gathers together a "packet" of characters of $G(k)$, including $\mathrm{R}_T^G\,\theta$, that each lift to $\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta}$.

Suppose we have a *Borel-torus pair* (B, T) for G; that is, we have a maximal k-torus $T \subseteq G$, together with a Borel subgroup $B \subseteq G$ containing T, not necessarily defined over k. From Proposition 1, we will obtain an ε-invariant Borel-torus pair $(\widetilde{B}, \widetilde{T})$ in $\widetilde{G}$, where $\widetilde{T} = C_{\widetilde{G}}(T)$ as above. Let $\langle\varepsilon\rangle$ be the group generated by ε. Recall that an element of a reductive group is *regular* if the connected part of its centralizer is a torus.

Theorem. *Suppose $\tilde{s} \in \widetilde{G}(k)$ belongs to an ε-invariant, maximal k-torus $\tilde{S}$ and $\mathcal{N}(\tilde{s})$ is regular in $\widetilde{G}$. Then*

$$(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_\varepsilon(\tilde{s}) = \Bigg(\sum_{\tilde{g} \in G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^\varepsilon} \mathrm{R}_{{}^{\tilde{g}}T}^{G}\,{}^{\tilde{g}}\theta\Bigg)(\mathcal{N}(\tilde{s})). \tag{0-1}$$

Remark. Let us comment on some of the terms in (0-1).

(a) Here is what we mean by $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_\varepsilon$. Since $\tilde{\theta}$ is an ε-invariant character of $\widetilde{T}(k)$, we have that ε acts on the Deligne–Lusztig variety corresponding to $(\widetilde{B}, \widetilde{T}, \tilde{\theta})$, and thus on the virtual representation whose character is $\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta}$. That is, even if this representation is reducible, we can form its ε-twisted character. To do so, extend $\tilde{\theta}$ to a character of $\widetilde{T}(k) \rtimes \langle\varepsilon\rangle$ by setting $\tilde{\theta}(\varepsilon) = 1$. Define the ε-twisted Deligne–Lusztig character $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\tilde{\theta})_\varepsilon$ induced from $\tilde{\theta}$ by $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_\varepsilon(g) = (\mathrm{R}_{\widetilde{T}\rtimes\langle\varepsilon\rangle}^{\widetilde{G}\rtimes\langle\varepsilon\rangle}\,\tilde{\theta})(g\varepsilon)$ for $g \in \widetilde{G}(k)$. (See [Digne and Michel 1994] for the definition of Deligne–Lusztig induction for nonconnected groups.)

Note that if ε is quasicentral, i.e., the Weyl groups $W(G,T)$ and $W(\widetilde{G},\widetilde{T})$ satisfy $W(G,T) = W(\widetilde{G},\widetilde{T})^{\varepsilon}$ (see [Digne and Michel 1994, Définition-Théorème 1.15] for equivalent formulations), then $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_{\varepsilon}$ doesn't depend on the choice of $\widetilde{B}$, from the remark after [Digne and Michel 1994, Théorème 4.5]. More generally, there could be several Borel subgroups $\widetilde{B} \subseteq \widetilde{G}$ such that $(\widetilde{B}^{\varepsilon})^{\circ} = B$, and we don't know if the twisted character is independent of the choice of $\widetilde{B}$. However, our theorem will remain valid for any such choice.

(b) On the right-hand side of (0-1), $\tilde{g}$ runs over a set of double coset representatives. For each such $\tilde{g}$, we have that ${}^{\tilde{g}}T$ is a maximal k-torus in G and ${}^{\tilde{g}}\theta$ is a complex-valued character of ${}^{\tilde{g}}T(k)$. The choice of representative $\tilde{g}$ affects the pair $({}^{\tilde{g}}T, {}^{\tilde{g}}\theta)$ only up to $G(k)$-conjugacy, so it does not affect the character $\mathrm{R}_{{}^{\tilde{g}}T}^{G}\,{}^{\tilde{g}}\theta$ appearing in the corresponding summand.

(c) We note that $G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}$ can be viewed as the set of $G(k)$-conjugacy classes of ε-invariant Borel-torus pairs that are $\tilde{G}(k)$-conjugate to $(\tilde{B},\widetilde{T})$. As $\tilde{g}$ runs over a set of double coset representatives, the k-tori ${}^{\tilde{g}}T \subset G$ are related in a way that is analogous to stable conjugacy, where $\widetilde{G}(k)$ plays the role usually played by $G(\bar{k})$, where $\bar{k}$ is an algebraic closure of k. Moreover, the set of characters appearing in the right-hand side of (0-1) is analogous to an L-packet, and their sum is "stable", in the sense that it is constant on the intersections with $G(k)$ of appropriate conjugacy classes in $\widetilde{G}(k)$.

Now let's consider some special cases.

(a) One can show that when ε is quasicentral, the index set for the summation is a singleton, so the theorem asserts that $(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_{\varepsilon}(\tilde{s}) = (\mathrm{R}_{T}^{G}\,\theta)(\mathcal{N}(\tilde{s}))$.

(b) In particular, suppose $\widetilde{G} = R_{E/k}G$ is obtained from G via restriction of scalars over a cyclic extension E/k and ε is the algebraic k-automorphism of $\widetilde{G}$ associated to the action of a generator of the Galois group $\mathrm{Gal}(E/k)$. Given a representation π of $G(k)$, one often has an associated representation $\tilde{\pi}$ of $\widetilde{G}(k)$, known as the *Shintani lift* of π. (See [Kawanaka 1987] for a discussion.) The character of π and the ε-twisted character of $\tilde{\pi}$ satisfy the Shintani relation $\Theta_{\tilde{\pi},\varepsilon} = \Theta_{\pi} \circ \mathcal{N}$, where $\mathcal{N}$ has been extended to a map on all (not necessarily semisimple) conjugacy classes. From work of Digne [1999, Corollaire 3.6], one already knows that if $\mathrm{R}_{T}^{G}\,\theta$ has a Shintani lift, then it must be $\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta}$. Thus, if one restricts attention to Deligne–Lusztig characters and to the kind of elements that we consider, our character relation is a generalization of Shintani's.

(c) Consider the automorphism of $\mathrm{GL}(2)$ given by $\varepsilon(g) = {}^{t}g^{-1}$. Then a relation analogous to that in our theorem holds for all irreducible characters, not just those of Deligne–Lusztig type. Moreover, the relation can be extended to unipotent elements, but it fails if $\tilde{s}$ is regular but $\mathcal{N}(\tilde{s})$ is singular.

Comparing our theorem with [Digne and Michel 1994, Proposition 2.12], we see the latter shows that the restriction to $G(k)$ of a twisted character of $\widetilde{G}(k)$ is a certain twisted character of $G(k)$, at least in the case where ε is quasicentral. On the other hand, our theorem concerns a lifting of characters from $G(k)$ to $\widetilde{G}(k)$ via the norm map, and this lifting is not an inverse of restriction. As expected, the two formulas agree in those situations where both are applicable, but such situations are rare. Moreover, while we make regularity assumptions on our element $\tilde{s}$, we do not assume that ε is quasicentral.

1. Preliminary results

Given any algebraic group H, we denote by H° the connected component of the identity in H. If S is an algebraic subgroup of H, then $C_H(S)$ denotes the centralizer of S in H. Similarly, for an element $h \in H$, we let $C_H(h)$ denote its centralizer.

As in Section 0, let k denote a finite field, $\tilde{G}$ a connected reductive k-group, and ε a semisimple k-automorphism of $\tilde{G}$ of finite order ℓ, and let $G = (\widetilde{G}^\varepsilon)^\circ$.

Most of the following result appears in the statement or proof of [Steinberg 1968, Theorem 8.2]. Other versions appear in [Kottwitz and Shelstad 1999, Theorem 1.1.A; Digne and Michel 1994, Théorème 1.8]. A version that includes the rationality of G is in [Adler and Lansky 2014, Proposition 3.5].

Proposition 1. *With notation as above, one has:*

- *G is a connected reductive k-group.*
- *For every ε-invariant Borel-torus pair $(\widetilde{B}, \widetilde{T})$ for $\widetilde{G}$, one has a Borel-torus pair $((\widetilde{B}^\varepsilon)^\circ, (\widetilde{T}^\varepsilon)^\circ)$ for G. Moreover, $(\widetilde{T}^\varepsilon)^\circ = \widetilde{T} \cap G$.*
- *For every Borel-torus pair (B, T) for G, one has an ε-invariant Borel-torus pair $(\widetilde{B}, \widetilde{T})$, where $\widetilde{T} = C_{\widetilde{G}}(T)$ and such that $(\widetilde{B}^\varepsilon)^\circ = B$.*

From now on, we fix a maximal k-torus $T \subseteq G$, and thus obtain an ε-invariant maximal k-torus $\widetilde{T} = C_{\widetilde{G}}(T)$ as in Proposition 1, and a norm map $\mathcal{N} : \widetilde{T} \to T$ as in Section 0.

The following result concerns conjugacy and twisted conjugacy in G and $\tilde{G}$.

Lemma 2. *Suppose $\widetilde{S}$ is an ε-invariant maximal k-torus in $\widetilde{G}$. Let $S = \tilde{S} \cap G$. Suppose that $\tilde{s} \in \widetilde{S}(k)$ and that $s := \mathcal{N}(\tilde{s}) \in S(k)$ is regular in $\widetilde{G}$. Let $\tilde{g} \in \widetilde{G}(k)$.*

(i) *There is an ε-invariant Borel subgroup of $\widetilde{G}$ containing $\widetilde{S}$.*

(ii) *If $\tilde{g}^{-1} s \tilde{g} \in T(k)$, then $\tilde{g}^{-1} \tilde{S} \tilde{g} = \widetilde{T}$.*

(iii) *If $\tilde{g}^{-1} \tilde{s} \tilde{g} \in \widetilde{T}(k)$ and $\tilde{g}^{-1} \varepsilon(\tilde{g}) \in \widetilde{T}(k)$, then $\tilde{g}^{-1} S \tilde{g} = T$.*

(iv) *If $\tilde{g}^{-1} \tilde{s} \varepsilon(\tilde{g}) \in \widetilde{T}(k)$, then $\tilde{g}^{-1} \varepsilon(\tilde{g}) \in \widetilde{T}(k)$.*

Proof. To prove (i), let S' be a maximal k-torus in G such that $s \in S'(k)$, and let $\widetilde{S}' = C_{\widetilde{G}}(S')$. From Proposition 1, we have that $\widetilde{S}'$ is a maximal torus in $\widetilde{G}$, and it is contained in an ε-invariant Borel subgroup of $\widetilde{G}$. But $\widetilde{S}' = C_{\widetilde{G}}(S') \subseteq C_{\widetilde{G}}(s)^\circ = \widetilde{S}$, and therefore, $\widetilde{S}' = \widetilde{S}$.

To prove (ii), note that $\tilde{g}^{-1}\tilde{S}\tilde{g} = \tilde{g}^{-1}C_{\widetilde{G}}(s)^\circ\tilde{g} = C_{\widetilde{G}}(\tilde{g}^{-1}s\tilde{g})^\circ = \widetilde{T}$.

For (iii), it follows immediately from the assumptions that $\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}) \in \widetilde{T}(k)$. Then $\tilde{g}^{-1}s\tilde{g} = \mathcal{N}(\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})) \in \mathcal{N}(\widetilde{T}(k)) \subseteq T(k)$. From (ii), $\tilde{g}^{-1}\tilde{S}\tilde{g} = \widetilde{T}$. From (i) and Proposition 1, it is thus enough to show that $(\tilde{g}^{-1}\tilde{S}\tilde{g})^\varepsilon = \tilde{g}^{-1}\tilde{S}^\varepsilon\tilde{g}$. For $x \in \widetilde{S}(\bar{k})$,

$$\begin{aligned}\varepsilon(\tilde{g}^{-1}x\tilde{g}) &= \varepsilon(\tilde{g})^{-1}\varepsilon(x)\varepsilon(\tilde{g}) \\ &= [\tilde{g}^{-1}\varepsilon(\tilde{g})]^{-1} \cdot \tilde{g}^{-1}\varepsilon(x)\tilde{g} \cdot [\tilde{g}^{-1}\varepsilon(\tilde{g})] \\ &= \tilde{g}^{-1}\varepsilon(x)\tilde{g},\end{aligned}$$

so $\varepsilon(x) = x$ if and only if $\varepsilon(\tilde{g}^{-1}x\tilde{g}) = \tilde{g}^{-1}x\tilde{g}$.

To prove (iv), note that as above, $\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}) \in \widetilde{T}(k)$ implies $\tilde{g}^{-1}s\tilde{g} \in T(k)$. By (ii), we have $\tilde{g}^{-1}\tilde{s}\tilde{g} \in \tilde{T}(k)$. Therefore,

$$\tilde{g}^{-1}\varepsilon(\tilde{g}) = (\tilde{g}^{-1}\tilde{s}\tilde{g})\ (\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}))^{-1} \in \widetilde{T}(k). \qquad \square$$

Now we consider a property of certain double coset spaces, whose proof is straightforward.

Lemma 3. *Let $\tilde{R}$ denote a set of representatives for the double coset space*

$$G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^\varepsilon.$$

Define the map $\phi : G(k) \times \tilde{R} \to (\widetilde{G}(k)/\widetilde{T}(k))^\varepsilon$ by $(g, \tilde{r}) \mapsto g\tilde{r}\widetilde{T}(k)$. Then

(i) *the map ϕ is surjective*;

(ii) *we have $\phi(g, \tilde{r}) = \phi(g', \tilde{r}')$ if and only if $\tilde{r} = \tilde{r}'$ and $g^{-1}g' \in {}^{\tilde{r}}T(k)$.*

Now we recall some facts about Deligne–Lusztig virtual characters.

Lemma 4. *One has the following*:

(i) *If $x \in G(k)$ is a regular element and θ is a complex character of $T(k)$, then*

$$(\mathrm{R}_T^G\,\theta)(x) = \sum_{\substack{g \in G(k)/T(k) \\ g^{-1}xg \in T(k)}} \theta(g^{-1}xg).$$

(ii) *Suppose that $\tilde{x} \in \tilde{G}(k)$ and that $\tilde{x}\varepsilon$ is a regular element of $\tilde{G}(k) \rtimes \langle\varepsilon\rangle$. Let $\tilde{\theta}$ be an ε-invariant character of $\widetilde{T}(k)$, extended trivially to $\widetilde{T}(k) \rtimes \langle\varepsilon\rangle$. Then*

$$(\mathrm{R}_{\widetilde{T}}^{\widetilde{G}}\,\tilde{\theta})_\varepsilon(\tilde{x}) = \frac{1}{|\widetilde{T}(k) \rtimes \langle\varepsilon\rangle|} \sum_{\substack{\tilde{h} \in \widetilde{G}(k) \rtimes \langle\varepsilon\rangle \\ \tilde{x}\varepsilon \in \tilde{h}(\widetilde{T}(k) \rtimes \langle\varepsilon\rangle)\tilde{h}^{-1}}} \tilde{\theta}(\tilde{h}^{-1}(\tilde{x}\varepsilon)\tilde{h}).$$

Proof. For each formula, see [Digne and Michel 1994, Proposition 2.6]. For the second formula, note that $C_{\widetilde{G}}(\tilde{x}\varepsilon)$ contains no nontrivial unipotent elements. Thus, if we let S denote its connected part, then S is a torus, and the Green function $Q_{S(k)}^{S(k)}$ that arises in this proposition takes the value $|S(k)|$ at $(1, 1)$. □

2. Proof of the main theorem

From Lemma 2(i), we know $\widetilde{S}$ is contained in an ε-invariant Borel subgroup of $\widetilde{G}$, so it follows from Proposition 1 that $S := \widetilde{S} \cap G$ is a maximal k-torus in G, and $\widetilde{S} = C_{\widetilde{G}}(S)$.

Since $(\tilde{s}\varepsilon)^{\ell} = \mathcal{N}(\tilde{s})$, which is assumed regular in $\widetilde{G}$ and thus in $\widetilde{G} \rtimes \langle\varepsilon\rangle$, we must have that $\tilde{s}\varepsilon$ is also regular in $\widetilde{G} \rtimes \langle\varepsilon\rangle$. Lemma 4(ii) then implies that the left-hand side of (0-1) (hereafter denoted LHS) is equal to

$$\frac{1}{\ell|\widetilde{T}(k)|}\sum \tilde{\theta}(\tilde{h}^{-1}(\tilde{s}\varepsilon)\tilde{h}) = \frac{1}{\ell|\widetilde{T}(k)|}\sum \tilde{\theta}(\tilde{h}^{-1}\cdot\tilde{s}\cdot\varepsilon(\tilde{h})\cdot\varepsilon),$$

where each sum is over the set

$$\bigl\{\tilde{h} \in \widetilde{G}(k) \rtimes \langle\varepsilon\rangle \bigm| \tilde{s}\varepsilon \in \tilde{h}(\widetilde{T}(k) \rtimes \langle\varepsilon\rangle)\tilde{h}^{-1}\bigr\}.$$

If $\tilde{g} \in \widetilde{G}(k)$, then $\tilde{g}\varepsilon^i$ belongs to the index set if and only if $\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}) \in \widetilde{T}(k)$. The set of such elements $\tilde{g}$ is a union of left cosets of $\widetilde{T}(k)$. Thus,

$$\begin{aligned}
\text{LHS} &= \frac{1}{\ell|\widetilde{T}(k)|}\sum_{i=0}^{\ell-1}\sum_{\substack{\tilde{g}\in\widetilde{G}(k)\\ \tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})\in\widetilde{T}(k)}} \tilde{\theta}\bigl((\tilde{g}\varepsilon^i)^{-1}\cdot\tilde{s}\cdot\varepsilon(\tilde{g}\varepsilon^i)\cdot\varepsilon\bigr)\\
&= \frac{1}{|\widetilde{T}(k)|}\sum_{\substack{\tilde{g}\in\widetilde{G}(k)\\ \tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})\in\widetilde{T}(k)}} \tilde{\theta}(\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}))\\
&= \sum_{\substack{\tilde{g}\in\widetilde{G}(k)/\widetilde{T}(k)\\ \tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})\in\widetilde{T}(k)}} \tilde{\theta}(\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g}))\\
&= \sum_{\substack{\tilde{g}\in(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}\\ \tilde{g}^{-1}\tilde{s}\tilde{g}\in\widetilde{T}(k)}} \tilde{\theta}(\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})),
\end{aligned}$$

where the last equality follows from Lemma 2(iv). Let $s = \mathcal{N}(\tilde{s})$ and note that $\mathcal{N}(\tilde{g}^{-1}\tilde{s}\varepsilon(\tilde{g})) = \tilde{g}^{-1}s\tilde{g}$ for $\tilde{g} \in \widetilde{G}(k)$. Thus LHS is equal to

$$\sum_{\substack{\tilde{g}\in(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}\\ \tilde{g}^{-1}\tilde{s}\tilde{g}\in\widetilde{T}(k)}} {}^{\tilde{g}}\theta(s). \tag{2-1}$$

On the other hand, it follows from Lemma 4(i) that the right-hand side of (0-1) is equal to

$$\begin{aligned}\sum_{\tilde{x}\in G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}} (\mathrm{R}^{G}_{\tilde{x}T}\,{}^{\tilde{x}}\theta)(s) &= \sum_{\tilde{x}\in G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}} \sum_{\substack{g\in G(k)/{}^{\tilde{x}}T(k)\\ g^{-1}sg\in{}^{\tilde{x}}T(k)}} {}^{\tilde{x}}\theta(g^{-1}sg)\\ &= \sum_{\tilde{x}\in G(k)\backslash(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}} \sum_{\substack{g\in G(k)/{}^{\tilde{x}}T(k)\\ (g\tilde{x})^{-1}s(g\tilde{x})\in T(k)}} {}^{g\tilde{x}}\theta(s)\\ &= \sum_{\substack{\tilde{g}\in(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}\\ \tilde{g}^{-1}s\tilde{g}\in T(k)}} {}^{\tilde{g}}\theta(s),\end{aligned}$$

where the final equality follows from Lemma 3. The last sum above is equal to (2-1) by Lemma 2(ii) and (iii), which together imply that for $\tilde{g}\in(\widetilde{G}(k)/\widetilde{T}(k))^{\varepsilon}$, we have $\tilde{g}^{-1}s\tilde{g}\in T(k)$ if and only if $\tilde{g}^{-1}\tilde{s}\tilde{g}\in\widetilde{T}(k)$.

Acknowledgements

The authors were partially supported by a National Science Foundation Focused Research Group grant (DMS-0854844). Morgan was partially supported by the Frank Cox Jones Scholarship Fund and a Dean's Undergraduate Research Award from the College of Arts and Sciences of American University. We have benefited from conversations with Prof. Jeffrey Hakim and from the comments of an anonymous referee.

References

[Adler and Lansky 2014] J. D. Adler and J. M. Lansky, "Lifting representations of finite reductive groups, I: Semisimple conjugacy classes", *Canad. J. Math.* **66**:6 (2014), 1201–1224. MR 3270781 Zbl 1309.20037

[Digne 1987] F. Digne, "Shintani descent and $\mathscr{L}$ functions on Deligne–Lusztig varieties", pp. 61–68 in *The Arcata Conference on Representations of Finite Groups* (Arcata, CA, 1986), edited by P. Fong, Proc. Sympos. Pure Math. **47**, Amer. Math. Soc., Providence, RI, 1987. MR 933350 Zbl 0657.20036

[Digne 1999] F. Digne, "Descente de Shintani et restriction des scalaires", *J. London Math. Soc.* (2) **59**:3 (1999), 867–880. MR 1709085 Zbl 0947.20025

[Digne and Michel 1994] F. Digne and J. Michel, "Groupes réductifs non connexes", *Ann. Sci. École Norm. Sup.* (4) **27**:3 (1994), 345–406. MR 1272294 Zbl 0846.20040

[Gyoja 1979] A. Gyoja, "Liftings of irreducible characters of finite reductive groups", *Osaka J. Math.* **16**:1 (1979), 1–30. MR 527011 Zbl 0416.20032

[Kawanaka 1987] N. Kawanaka, "Shintani lifting and Gel'fand–Graev representations", pp. 147–163 in *The Arcata Conference on Representations of Finite Groups* (Arcata, CA, 1986), edited by P. Fong, Proc. Sympos. Pure Math. **47**, Amer. Math. Soc., Providence, RI, 1987. MR 933357 Zbl 0654.20046

[Kottwitz and Shelstad 1999] R. E. Kottwitz and D. Shelstad, *Foundations of twisted endoscopy*, Astérisque **255**, Société Mathématique de France, Paris, 1999. MR 1687096 Zbl 0958.22013

[Lusztig 1984] G. Lusztig, *Characters of reductive groups over a finite field*, Annals of Mathematics Studies **107**, Princeton University Press, 1984. MR 742472 Zbl 0556.20033

[Silberger and Zink 2005] A. J. Silberger and E.-W. Zink, "An explicit matching theorem for level zero discrete series of unit groups of $\mathfrak{p}$-adic simple algebras", *J. Reine Angew. Math.* **585** (2005), 173–235. MR 2164626 Zbl 1087.22012

[Steinberg 1968] R. Steinberg, *Endomorphisms of linear algebraic groups*, Memoirs of the American Mathematical Society **80**, American Mathematical Society, Providence, R.I., 1968. MR 0230728 Zbl 0164.02902

Received: 2015-05-27 Revised: 2015-11-01 Accepted: 2015-11-03

jadler@american.edu — *Department of Mathematics and Statistics, American University, Washington, DC 20016, United States*

m.cassel@columbia.edu — *Columbia University, New York, NY 10027, United States*

lansky@american.edu — *Department of Mathematics and Statistics, American University, Washington, DC 20016, United States*

emma.morgan@tufts.edu — *Office of Institutional Research and Evaluation, Tufts University, Medford, MA 02155, United States*

yifei@math.harvard.edu — *Department of Mathematics, Harvard University, Cambridge, MA 02138, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.813

Spectrum of a composition operator with automorphic symbol

Robert F. Allen, Thong M. Le and Matthew A. Pons

(Communicated by Stephan Garcia)

We give a complete characterization of the spectrum of composition operators, induced by an automorphism of the open unit disk, acting on a family of Banach spaces of analytic functions that includes the Bloch space and BMOA. We show that for parabolic and hyperbolic automorphisms the spectrum is the unit circle. For the case of elliptic automorphisms, the spectrum is either the unit circle or a finite cyclic subgroup of the unit circle.

1. Introduction

For an analytic self-map φ of the open unit disk $\mathbb{D}$ and a Banach space X of functions analytic on $\mathbb{D}$, we define the composition operator with symbol φ, denoted C_φ, by the rule $C_\varphi f = f \circ \varphi$ for all $f \in X$. The study of composition operators began formally with Nordgren's paper [1968], where he explored properties of composition operators acting on the Hardy Hilbert space H^2. Since then the study has proved to be an active area of research, most likely due to the fact that the study of such operators lies at the intersection of complex function theory and operator theory.

The spectrum of C_φ has been studied on many classical spaces of analytic functions, such as the Hardy spaces, Bergman spaces, weighted Hardy and Bergman spaces, Besov spaces, and the Dirichlet space. The interested reader is directed to [Cowen and MacCluer 1995] for general references.

The motivation for this paper was to determine the spectrum of a composition operator, induced by a disk automorphism, acting on the Bloch space. The Bloch space is the largest space of analytic functions on $\mathbb{D}$ that is Möbius invariant. This is one reason the Bloch space is a welcoming environment to study composition operators. The techniques developed apply to a larger class of spaces that includes the Bloch space.

The purpose of this paper is to determine the spectrum of C_φ acting on a family of Banach spaces, where φ is a disk automorphism. The spectrum will depend on

MSC2010: primary 47A10, 47B33; secondary 30H05.
Keywords: composition operator, spectrum, automorphism.

the fixed point classification of the automorphisms of $\mathbb{D}$. This is a standard approach to the study of composition operators induced by automorphisms. We show the spectrum of C_φ, acting on a particular family of Banach spaces, induced by a disk automorphism, must be a subset of the unit circle $\partial\mathbb{D}$, and in some instances is the entire unit circle. Finally, we compare these results to particular examples of classical spaces.

2. Preliminaries

2A. *Automorphisms.* The automorphisms of the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ are precisely the analytic bijections on $\mathbb{D}$ which have the form

$$\varphi(z) = \lambda \frac{a - z}{1 - \overline{a}z},$$

where λ is a unimodular constant and a is a point in $\mathbb{D}$. These automorphisms form a group under composition denoted by $\mathrm{Aut}(\mathbb{D})$. Every element of $\mathrm{Aut}(\mathbb{D})$ has two fixed points (counting multiplicity), and thus can be classified by the location of the fixed points:

elliptic: one fixed point in $\mathbb{D}$ and one in the complement of $\overline{\mathbb{D}}$;

parabolic: one fixed point on the unit circle $\partial\mathbb{D}$ (of multiplicity 2);

hyperbolic: two distinct fixed points on $\partial\mathbb{D}$.

Two disk automorphisms φ and ψ are conformally equivalent if there exists a disk automorphism τ for which $\psi = \tau \circ \varphi \circ \tau^{-1}$. Many properties of automorphisms are preserved under conformal equivalence. The main advantage of conformal equivalence is in the placement of the fixed points. Every elliptic disk automorphism is conformally equivalent to one whose fixed point in $\mathbb{D}$ is the origin.

Lemma 2.1. *Let φ be an elliptic disk automorphism with fixed point a in $\mathbb{D}$. Then φ is conformally equivalent to $\psi(z) = \lambda z$ where $\lambda = \varphi'(a)$.*

Proof. Let τ_a be the involution automorphism which interchanges 0 and a, that is

$$\tau_a(z) = \frac{a - z}{1 - \overline{a}z}.$$

Define $\psi = \tau_a \circ \varphi \circ \tau_a^{-1}$ on $\mathbb{D}$. Since a is a fixed point of φ, ψ fixes the origin, and is a rotation. So there is a unimodular constant λ such that $\psi(z) = \lambda z$. To complete the proof, we will show $\lambda = \varphi'(a)$. Observe $\psi'(z) = \lambda$ for all $z \in \mathbb{D}$. In particular

$$\lambda = \psi'(0) = \tau_a'\big(\varphi(\tau_a(0))\big)\varphi'\big(\tau_a(0)\big)\tau_a'(0) = \varphi'(a)\tau_a'(a)\tau_a'(0) = \varphi'(a).$$

Thus φ is conformally equivalent to the rotation $\psi(z) = \varphi'(a)z$. $\square$

Every parabolic disk automorphism is conformally equivalent to one whose fixed point (of multiplicity 2) is 1. The following lemma is found as Exercise 2.3.5c of [Cowen and MacCluer 1995], and a complete proof can be found in [Pons 2007].

Lemma 2.2 [Pons 2007, Lemma 4.1.2]. *Let φ be a parabolic disk automorphism. Then φ is conformally equivalent to either*

$$\psi_1(z)=\frac{(1+i)z-1}{z+i-1} \quad \textit{or} \quad \psi_2(z)=\frac{(1-i)z-1}{z-i-1}.$$

Every hyperbolic disk automorphism is conformally equivalent to one whose fixed points in $\partial\mathbb{D}$ are ± 1.

Lemma 2.3 [Nordgren 1968, Theorem 6]. *Let φ be a hyperbolic disk automorphism. Then, for some $r \in (0, 1)$, φ is conformally equivalent to*

$$\psi(z)=\frac{z+r}{1+rz}.$$

2B. ***The space of bounded analytic functions.*** The set of analytic functions on $\mathbb{D}$ is denoted by $H(\mathbb{D})$. The space of bounded analytic functions on $\mathbb{D}$, denoted $H^\infty = H^\infty(\mathbb{D})$, is a Banach space under the norm

$$\|f\|_\infty=\sup_{z\in\mathbb{D}}|f(z)|.$$

The bounded analytic functions on $\mathbb{D}$ is a rich space containing many interesting types of functions, such as polynomials and Blaschke products. In addition, the disk algebra $\mathcal{A}(\mathbb{D})$, the set of analytic functions on $\mathbb{D}$ continuous to $\partial\mathbb{D}$, is a closed subspace of H^∞.

The following two families of functions will be used in the next section. To prove these functions are in H^∞, we take a geometric approach using conformal mappings of the plane. To this effect let $\mathbb{H}_\ell$ and $\mathbb{H}_r$ denote the open left and right half planes respectively, i.e., $\mathbb{H}_\ell = \{\operatorname{Re} z < 0\}$ and $\mathbb{H}_r = \{\operatorname{Re} z > 0\}$.

Lemma 2.4. *For $s \geq 0$, the function*

$$f_s(z)=\exp\frac{s(z+1)}{z-1}$$

is in H^∞.

Proof. If $s = 0$, then f_s is identically 1. So, $f_s(z)$ is in H^∞. Now suppose $s > 0$. The function f_s is comprised of the functions

(1) $z \mapsto (z+1)/(z-1)$ (mapping $\mathbb{D}$ onto $\mathbb{H}_\ell$),

(2) $z \mapsto sz$ (mapping $\mathbb{H}_\ell$ onto $\mathbb{H}_\ell$),

(3) $z \mapsto e^z$ (mapping $\mathbb{H}_\ell$ onto $\mathbb{D}\setminus\{0\}$).

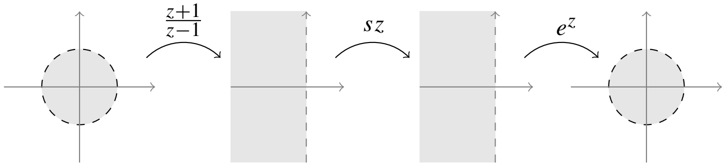


Figure 1. Map $f_s(z) = \exp \dfrac{s(z+1)}{z-1}$ for $s > 0$.

So f_s maps $\mathbb{D}$ into $\mathbb{D}$, as in Figure 1, and thus $f_s(z)$ is an element of H^∞. $\square$

Lemma 2.5. *For real value t, the function*

$$f_t(z) = \left(\frac{1+z}{1-z}\right)^{it}$$

is in H^∞.

Proof. For $t = 0$, f_t is identically 1, and thus is in H^∞. Now suppose $t > 0$. We will rewrite the function f_t as

$$f_t(z) = \exp\left(it \log \frac{1+z}{1-z}\right),$$

where log is the principle branch of the logarithm. Then f_t is comprised of the functions

(1) $z \mapsto (1+z)/(1-z)$ (mapping $\mathbb{D}$ onto $\mathbb{H}_r$),

(2) $z \mapsto \log z$ (mapping H_r onto the horizontal strip $S_h = \{0 < \operatorname{Im} z < 2\pi\}$),

(3) $z \mapsto itz$ (mapping S_h onto the vertical strip $S_v = \{-2\pi < \operatorname{Re} z < 0\}$),

(4) $z \mapsto e^z$ (mapping S_v into $A(e^{-2\pi}, 1) = \{e^{-2\pi} < |z| < 1\}$).

So f_t maps $\mathbb{D}$ into $A(e^{-2\pi}, 1) \subseteq \mathbb{D}$, as depicted in Figure 2. In the case of $t < 0$, the vertical strip S_v becomes $\{0 < \operatorname{Re} z < 2\pi\}$. The map $z \mapsto e^z$ takes S_v into $A(1, e^{2\pi}) \subseteq e^{2\pi}\mathbb{D}$, as depicted in Figure 3. In either case, $f_t(z)$ is an element of H^∞ since $\|f_t\|_\infty < e^{2\pi}$ for all $t \in \mathbb{R}$. $\square$

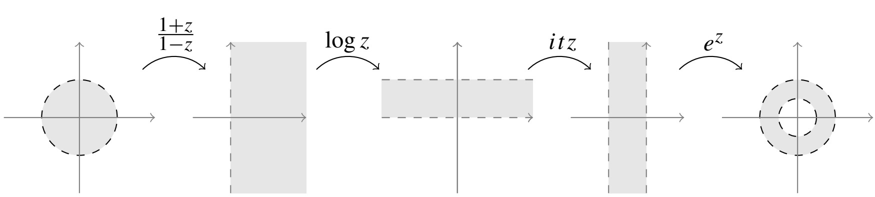


Figure 2. Map $f_t(z) = \exp\left(it \log \dfrac{1+z}{1-z}\right)$ for $t > 0$.

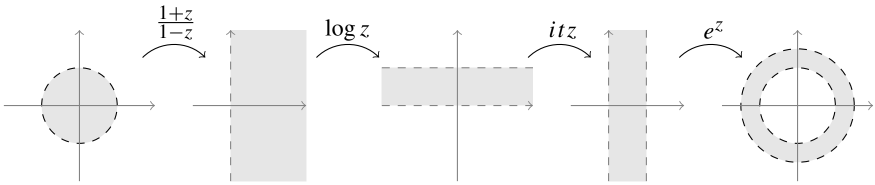


Figure 3. Map $f_t(z) = \exp\left(it \log \dfrac{1+z}{1-z}\right)$ for $t < 0$.

These functions above, together with the monomials, play such a pivotal role in Section 3 that we denote the union of these functions by $\mathcal{F}$, i.e.,

$$\mathcal{F} = \{f_s : s \geq 0\} \cup \{f_t : t \in \mathbb{R}\} \cup \{z^k : k \in \mathbb{N}\}.$$

2C. ***Spectrum of C_φ.*** In this section we collect useful results regarding the spectrum of operators on Banach spaces. For a bounded linear operator T on a Banach space X, the spectrum of T is given by

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ is not invertible}\}$$

where I denotes the identity operator on X. The spectrum is a nonempty, closed subset of $\mathbb{C}$. The spectral radius of T is given by

$$\rho(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\}.$$

Due to the fact that the spectrum is closed, we have the spectrum of T is contained in the closed disk centered at the origin of radius $\rho(T)$.

Determining the spectrum of a particular composition operator can be difficult depending on the symbol of the operator and the space on which it is acting. However, the difficulties can be avoided if the operator is similar to a "simpler" operator. Linear operators S and T (not necessarily bounded) on a Banach space X are similar if there exists a bounded linear operator U on X, having bounded inverse, such that $T = USU^{-1}$. If S and T are both bounded operators, then similarity preserves the spectrum.

Theorem 2.6. *Let S and T be bounded operators on a Banach space X. If S and T are similar, then $\sigma(S) = \sigma(T)$.*

Proof. Suppose S and T are similar operators on X. By definition, there exists an invertible, bounded operator U such that $T = USU^{-1}$. Let $\lambda \in \mathbb{C}$ and observe that

$$\begin{aligned} T - \lambda I &= USU^{-1} - \lambda I \\ &= USU^{-1} - \lambda UU^{-1} \\ &= USU^{-1} - U(\lambda I)U^{-1} \\ &= U(S - \lambda I)U^{-1}. \end{aligned}$$

Thus, we have that $S-\lambda I$ is not invertible if and only if $T-\lambda I$ is not invertible. Therefore $\sigma(S)=\sigma(T)$. ☐

3. Main results

In this section, we determine the spectrum of C_φ for φ a disk automorphism acting on a particular family of Banach spaces of analytic functions. The spaces we consider will be denoted by $\mathcal{X}$ and have the following properties:

(i) $\mathcal{X}$ contains $\mathcal{F}$,

(ii) for all $\varphi \in \mathrm{Aut}(\mathbb{D})$, C_φ is bounded on $\mathcal{X}$ and $\rho(C_\varphi)=1$.

The set of automorphisms of $\mathbb{D}$, as seen previously, is a very nice subset of the analytic self-maps of $\mathbb{D}$. By property (ii), every composition operator induced by a disk automorphism is bounded on $\mathcal{X}$. In fact, every such composition operator is invertible. This result, that we prove below, can be viewed as a consequence of Theorem 1.6 of [Cowen and MacCluer 1995].

Proposition 3.1. *Let φ be a disk automorphism and C_φ the induced composition operator on $\mathcal{X}$. Then C_φ is invertible with inverse $C_\varphi^{-1}=C_{\varphi^{-1}}$.*

Proof. Since $\varphi \in \mathrm{Aut}(\mathbb{D})$, φ is invertible, and φ^{-1} is an automorphism. The composition operator $C_{\varphi^{-1}}$ is bounded by property (ii) and

$$C_\varphi(C_{\varphi^{-1}}(f)) = C_\varphi(f\circ\varphi^{-1}) = f\circ\varphi^{-1}\circ\varphi = f,$$
$$C_{\varphi^{-1}}(C_\varphi(f)) = C_{\varphi^{-1}}(f\circ\varphi) = f\circ\varphi\circ\varphi^{-1} = f\ .$$

Therefore, C_φ is invertible with $C_\varphi^{-1}=C_{\varphi^{-1}}$. ☐

Since the spectral radius of C_φ on $\mathcal{X}$ is 1 for $\varphi\in\mathrm{Aut}(\mathbb{D})$, we see that the search for the spectrum can be restricted to subsets of $\overline{\mathbb{D}}$. However, our search can be refined further to subsets of the unit circle.

Theorem 3.2. *Let φ be a disk automorphism and C_φ the induced composition operator on $\mathcal{X}$. Then $\sigma(C_\varphi)\subseteq\partial\mathbb{D}$.*

Proof. By property (ii) of $\mathcal{X}$, we have $\rho(C_\varphi)=1$. So, $\sigma(C_\varphi)\subseteq\overline{\mathbb{D}}$. Since, by Proposition 3.1, C_φ is invertible with the inverse $C_\varphi^{-1}=C_{\varphi^{-1}}$, then $0\notin\sigma(C_\varphi)$. So, the function $f(z)=z^{-1}$ is analytic in some neighborhood of $\sigma(C_\varphi)$. By the Spectral Mapping Theorem (see Theorem 5.14 of [MacCluer 2009]), we have $\sigma(f\circ C_\varphi)=f(\sigma(C_\varphi))$, and so,

$$\sigma(C_{\varphi^{-1}})=\sigma(C_\varphi^{-1})=\sigma(C_\varphi)^{-1}=\{\lambda^{-1}:\lambda\in\sigma(C_\varphi)\}\ .$$

Since $\varphi^{-1}\in\mathrm{Aut}(\mathbb{D})$, $\sigma(C_{\varphi^{-1}})\subseteq\overline{\mathbb{D}}$. Thus for $\lambda\in\sigma(C_\varphi)$, both λ and λ^{-1} are in $\overline{\mathbb{D}}$. This implies $\lambda\in\partial\mathbb{D}$. So $\sigma(C_\varphi)\subseteq\partial\mathbb{D}$, as desired. ☐

Since the disk automorphisms are classified into three categories, according to fixed points, we will treat each type of automorphism separately. However, the strategy to determine $\sigma(C_\varphi)$ is the same. For a disk automorphism φ, we have shown φ to be conformally equivalent to a particularly "nice" disk automorphism: in the elliptic case a disk automorphism that fixes 0, in the parabolic case a disk automorphism that fixes 1, and in the hyperbolic case a disk automorphism that fixes ± 1. In the next result, we show that conformally equivalent automorphisms induce similar composition operators on $\mathcal{X}$. This result is not unique to the space $\mathcal{X}$, but is true for any space for which automorphisms induce bounded composition operators (see p. 250 of [Cowen and MacCluer 1995]).

Proposition 3.3. *Let φ and ψ be conformally equivalent disk automorphisms. Then the induced composition operators C_φ and C_ψ on $\mathcal{X}$ are similar.*

Proof. Suppose φ and ψ are conformally equivalent disk automorphisms. Then there exists a disk automorphism τ such that $\psi = \tau \circ \varphi \circ \tau^{-1}$. For $f \in \mathcal{X}$, observe

$$C_\psi f = f \circ (\tau \circ \varphi \circ \tau^{-1}) = ((f \circ \tau) \circ \varphi) \circ \tau^{-1} = (C_{\tau^{-1}} C_\varphi C_\tau) f.$$

Since $C_{\tau^{-1}}$ is bounded and invertible on $\mathcal{X}$ with $C_{\tau^{-1}}^{-1} = C_\tau$, then $C_\psi = C_{\tau^{-1}} C_\varphi C_{\tau^{-1}}^{-1}$. Therefore C_φ and C_ψ are similar. □

With Proposition 3.3 and Lemmas 2.1, 2.2, and 2.3, it suffices to determine the spectrum of composition operators induced by these "nice" disk automorphisms, since similarity of bounded operators preserves the spectrum.

Theorem 3.4. *Let φ be an elliptic disk automorphism with fixed point a in $\mathbb{D}$. Then the spectrum of C_φ acting on $\mathcal{X}$ is the closure of the positive powers of $\varphi'(a)$. Moreover, this closure is a finite subgroup of the unit circle if $\varphi'(a)^n = 1$ for some natural number n, and is the unit circle otherwise.*

Proof. By Lemma 2.1, φ is conformally equivalent to $\psi(z) = \lambda z$ where $\lambda = \varphi'(a)$. By Proposition 3.3, it suffices to show that $\sigma(C_\psi)$ is the closure of the positive powers of λ. Let $G = \langle \lambda \rangle = \{\lambda^k : k \in \mathbb{N}\}$, which is a subset of $\partial\mathbb{D}$ since $|\lambda| = 1$. For each $k \in \mathbb{N}$, the function $f_k(z) = z^k$ is in $\mathcal{X}$ by property (i), and we have $(C_\psi f_k)(z) = \lambda^k f_k(z)$. Thus λ^k is an eigenvalue of C_ψ corresponding to the eigenfunction f_k. So $G \subseteq \sigma(C_\psi)$, and since the spectrum is closed, we have $\overline{G} \subseteq \sigma(C_\psi) = \sigma(C_\varphi)$. If the order of λ is infinite, then G is dense in $\partial\mathbb{D}$, and so $\overline{G} = \partial\mathbb{D}$.

Now suppose λ has order $m < \infty$. Then $G = \{\lambda^k : k = 1, \dots, m\}$. So, $\overline{G} = G$. We now wish to show $\sigma(C_\psi) \subseteq \overline{G}$. Since $\sigma(C_\psi) \subseteq \partial\mathbb{D}$ by Theorem 3.2 it suffices to show that if $\mu \in \partial\mathbb{D} \setminus \overline{G}$ then $\mu \notin \sigma(C_\psi)$. Suppose $\mu \in \partial\mathbb{D} \setminus \overline{G}$.

Since $\mu \notin \overline{G}$, it clear that $\mu \notin G$ and $\mu^m \neq 1$. In order to show $\mu \notin \sigma(C_\psi)$, we will show that $C_\psi - \mu I$ is invertible by proving that for every $g \in \mathcal{X}$, there exists a unique $f \in \mathcal{X}$ such that $f \circ \psi - \mu f = g$.

Since the order of λ is m, we have

$$\psi^{(m)}(z) = (\underbrace{\psi \circ \cdots \circ \psi}_{m-\text{times}})(z) = \lambda^m z = z.$$

By repeated composition with ψ, we obtain the system of linear equations:

$$\begin{aligned} f \circ \psi - \mu f &= g \\ f \circ \psi^{(2)} - \mu(f \circ \psi) &= g \circ \psi \\ &\vdots \\ f - \mu(f \circ \psi^{(m-1)}) &= g \circ \psi^{(m-1)} . \end{aligned}$$

This system of linear equations can be expressed as the matrix equation $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} -\mu & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\mu & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \vdots & & \ddots & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 & -\mu \end{bmatrix}, \vec{x} = \begin{bmatrix} f \\ f \circ \psi \\ \vdots \\ \vdots \\ f \circ \psi^{(m-2)} \\ f \circ \psi^{(m-1)} \end{bmatrix}, \text{ and } \vec{b} = \begin{bmatrix} g \\ g \circ \psi \\ \vdots \\ \vdots \\ g \circ \psi^{(m-2)} \\ g \circ \psi^{(m-1)} \end{bmatrix}.$$

The determinant of A is $(-1)^m(\mu^m - 1)$, which is not zero since $\mu \notin G$. Thus there is a unique solution for $\vec{x}$. It gives us the unique solution f, which is a finite linear combination of function in $\mathcal{X}$ of the form $g \circ \psi^{(j-1)}$ for $j = 1, \ldots, m$, and thus f is in $\mathcal{X}$. It follows that $C_\psi - \mu I$ is invertible. So, $\mu \notin \sigma(C_\psi)$. Therefore, $\sigma(C_\varphi) = \sigma(C_\psi) \subseteq \overline{G}$. □

Theorem 3.5. *Let φ be a parabolic disk automorphism. Then the spectrum of C_φ acting on $\mathcal{X}$ is the unit circle.*

Proof. From Lemma 2.2, φ is conformally equivalent to either

$$\psi_1(z) = \frac{(1+i)z - 1}{z + i - 1} \quad \text{or} \quad \psi_2(z) = \frac{(1-i)z - 1}{z - i - 1}.$$

By Theorems 3.2 and 3.3 it suffices to show that $\partial\mathbb{D}$ is a subset of $\sigma(C_{\psi_1})$ and $\sigma(C_{\psi_2})$.

First suppose φ is conformally equivalent to ψ_1. Consider the function

$$f_s(z) = \exp \frac{s(z+1)}{z-1}$$

for $s \geq 0$. By property (i), f_s is in $\mathcal{X}$. Observe

$$\begin{aligned}(C_{\psi_1} f_s)(z) = f_s(\psi_1(z)) &= f_s\left(\frac{(1+i)z-1}{z+i-1}\right) \\ &= \exp \frac{s\left(\frac{(1+i)z-1}{z+i-1}+1\right)}{\frac{(1+i)z-1}{z+i-1}-1} = \exp \frac{s((1+i)z-1+z+i-1)}{(1+i)z-1-z-i+1} \\ &= \exp \frac{s((2+i)z+i-2)}{i(z-1)} = \exp \frac{s((1-2i)z+1+2i)}{z-1} \\ &= \exp\left(\frac{s(z+1)}{z-1} - 2is\right) = e^{i(-2s)} f_s(z).\end{aligned}$$

So, f_s is an eigenfunction of C_{ψ_1} for $s \geq 0$. Then, $\partial\mathbb{D} = \{e^{i(-2s)} : s \geq 0\}$ is a subset of $\sigma(C_{\psi_1})$. If φ is conformally equivalent to ψ_2, then by a similar calculation, we have

$$(C_{\psi_2} f_s)(z) = e^{2is} f_s(z),$$

and so $\partial\mathbb{D} = \{e^{2is} : s \geq 0\}$ is a subset of $\sigma(C_{\psi_2})$. Therefore, $\sigma(C_\varphi) = \partial\mathbb{D}$, as desired. $\square$

Theorem 3.6. *Let φ be a hyperbolic disk automorphism. Then the spectrum of C_φ acting on $\mathcal{X}$ is the unit circle.*

Proof. From Lemma 2.3, φ is conformally equivalent to $\psi(z) = (z+r)/(1+rz)$ for some $r \in (0, 1)$. By Theorems 3.2 and 3.3 it suffices to show that $\partial\mathbb{D} \subseteq \sigma(C_\psi)$. Consider the function

$$f_t(z) = \left(\frac{1+z}{1-z}\right)^{it}$$

for $t \in \mathbb{R}$. By property (i), f_t is in $\mathcal{X}$. Observe

$$\begin{aligned}(C_\psi f_t)(z) = f_t(\psi(z)) &= f_t\left(\frac{z+r}{1+rz}\right) \\ &= \left(\frac{1+\frac{z+r}{1+rz}}{1-\frac{z+r}{1+rz}}\right)^{it} = \left(\frac{1+rz+z+r}{1+rz-z-r}\right)^{it} \\ &= \left(\frac{(r+1)z+(r+1)}{(r-1)z-(r-1)}\right)^{it} \\ &= \left(\frac{r+1}{r-1}\right)^{it} f_t(z).\end{aligned}$$

So, f_t is an eigenfunction of C_ψ for t real. Then $\partial\mathbb{D} = \left\{\left((r+1)/(r-1)\right)^{it} : 0 < r < 1, t \in \mathbb{R}\right\}$ is a subset of $\sigma(C_\psi) = \sigma(C_\varphi)$. Therefore $\sigma(C_\varphi) = \partial\mathbb{D}$, as desired. $\square$

4. Examples and comparisons

In this section we first consider examples of spaces that satisfy the properties of $\mathcal{X}$. For these spaces, our results characterize the spectrum of composition operators induced by disk automorphisms. Lastly, we consider spaces that do not satisfy the properties of $\mathcal{X}$ but for which the spectrum of composition operators induced by automorphisms is known. We will compare the spectra for those spaces with the characterization for $\mathcal{X}$.

4A. ***Examples.*** First, we will discuss examples of spaces that satisfy the properties of $\mathcal{X}$.

4A1. *Bounded analytic functions.* The property (i) of $\mathcal{X}$ is satisfied by H^∞ by Lemmas 2.4 and 2.5. In fact, on H^∞, any analytic self-map of $\mathbb{D}$ induces a bounded composition operator C_φ such that $\|C_\varphi\| = 1$. Equality is achieved since H^∞ contains the constant function 1. The spectral radius formula (see Theorem 5.15 of [MacCluer 2009]) then implies that $\rho(C_\varphi) = 1$. Thus, property (ii) is satisfied. Thus H^∞ belongs to the family of Banach spaces of analytic functions $\mathcal{X}$.

4A2. *Bloch space.* The Bloch space on $\mathbb{D}$, denoted $\mathcal{B} = \mathcal{B}(\mathbb{D})$, is the space of analytic functions on $\mathbb{D}$ such that $\beta_f = \sup_{z\in\mathbb{D}} (1-|z|^2)|f'(z)| < \infty$. The quantity β_f is a seminorm, called the Bloch seminorm. The Bloch space is a Banach space under the norm

$$\|f\|_{\mathcal{B}} = |f(0)| + \beta_f.$$

It is well-known that $\mathcal{B}$ is a Banach space of analytic functions that contains H^∞, and thus satisfies property (i) of $\mathcal{X}$. In fact, every analytic self-map of $\mathbb{D}$ induces a bounded composition operator on $\mathcal{B}$ (see [Arazy et al. 1985, p. 126]). Donaway, in his Ph.D. thesis, proved the spectral radius of every composition operator induced by an analytic function on $\mathbb{D}$, and in particular the disk automorphisms, is 1, by [Donaway 1999, Corollary 3.9]. So the Bloch space satisfies all the properties of $\mathcal{X}$.

4A3. *Analytic functions of bounded mean oscillation.* The space of analytic functions on $\mathbb{D}$ with bounded mean oscillation on $\partial\mathbb{D}$, denoted BMOA, is defined to be the set of functions in $H(\mathbb{D})$ such that

$$\|f\|_* = \sup_{z\in\mathbb{D}} \|f\circ\tau_a - f(z)\|_{H^2} < \infty,$$

where H^2 is defined in Section 4B1. The space BMOA is a Banach space under the norm

$$\|f\|_{\mathrm{BMOA}} = |f(0)| + \|f\|_*.$$

It is well-known that BMOA is a Banach space of analytic functions, a subspace of the Bloch space, and contains H^∞ as a subspace since $\|f\|_{\mathrm{BMOA}} \le 3\|f\|_\infty$.

Thus property (i) is satisfied by BMOA. The following result shows property (ii) is satisfied by BMOA also.

Theorem 4.1. *Let φ be an analytic self-map of $\mathbb{D}$. Then C_φ acting on* BMOA *is bounded and $\rho(C_\varphi) = 1$.*

Proof. As a result of the Littlewood subordination principle (see Theorem 1.7 of [Duren 1970]), every analytic self-map φ of $\mathbb{D}$ induces a bounded composition operator on BMOA.

To compute the spectral radius of C_φ acting on BMOA, we first estimate the norm. By Corollary 2.2 of [Laitila 2009] there is a constant $M > 0$, independent of φ, such that

$$\|C_\varphi\| \le M\left(\sup_{a\in\mathbb{D}}\|\tau_{\varphi(a)} \circ \varphi \circ \tau_a\|_{H^2} + \log\frac{2}{1-|\varphi(0)|^2}\right). \tag{4-1}$$

Since the function $\tau_{\varphi(a)} \circ \varphi \circ \tau_a$ is a composition of self-maps of the disk, the first term on the right is bounded above by 1. Also,

$$\frac{1}{1-|\varphi(0)|^2} \le \frac{1+|\varphi(0)|}{1-|\varphi(0)|} \le \frac{2}{1-|\varphi(0)|}$$

and hence

$$\log\left(\frac{2}{1-|\varphi(0)|^2}\right) \le \log\left(\frac{4}{1-|\varphi(0)|}\right) \le 2\log 2 - \log(1-|\varphi(0)|).$$

Applying these estimates to Equation (4-1), we have

$$\|C_\varphi\| \le M(1+2\log 2) - M\log(1-|\varphi(0)|).$$

This immediately implies that

$$\|C_{\varphi_n}\| \le M(1+2\log 2) - M\log(1-|\varphi_n(0)|)$$

and it follows that $\rho(C_\varphi) = 1$ for all bounded composition operators acting on BMOA by Theorem 3.7 of [Donaway 1999]. □

Thus BMOA satisfies all the properties of $\mathcal{X}$.

4B. ***Comparisons.*** We now investigate spaces that do not satisfy the properties of $\mathcal{X}$. We compare the spectrum of induced composition operators on these spaces with those on $\mathcal{X}$.

4B1. *Hardy spaces.* For $1 \le p < \infty$, the Hardy space, denoted $H^p = H^p(\mathbb{D})$, is the space of analytic functions on $\mathbb{D}$ such that

$$\|f\|_{H^p}^p = \sup_{0<r<1}\int_{\mathbb{D}}\left|f(re^{i\theta})\right|^p\frac{d\theta}{2\pi} < \infty.$$

Under this norm, the Hardy spaces are Banach spaces and for $p = 2$ it is a Hilbert space.

It is well known that H^p is a Banach space of analytic functions that contains H^∞ as a subspace. For the cases of an elliptic or parabolic automorphism φ, it is the case that $\rho(C_\varphi) = 1$ and the spectrum of C_φ on H^p is the same as for C_φ acting on $\mathcal{X}$ (see Theorem 3.9 of [Cowen and MacCluer 1995]). However, it is not the case that the spectral radius is 1 for every composition operator induced by an automorphism. In fact, if φ is hyperbolic, then $\rho(C_\varphi) = \varphi'(a)^{-1/p}$ where a is the Denjoy–Wolff point of φ (see Theorem 3.9 of [Cowen and MacCluer 1995]). In this situation, $\varphi'(a) < 1$ thus making $\rho(C_\varphi) > 1$. In turn, the spectrum is the annulus $\varphi'(a)^{1/p} \le |z| \le \varphi'(a)^{-1/p}$ (see Theorem 4.9 of [Hyvärinen et al. 2013]).

4B2. *Weighted Bergman spaces.* For $1 \le p < \infty$ and $\alpha > -1$, the standard weighted Bergman space, denoted $A^p_\alpha = A^p_\alpha(\mathbb{D})$, is the space of analytic functions on $\mathbb{D}$ such that

$$\|f\|^p_{A^p_\alpha} = \int_{\mathbb{D}} \bigl(1 - |z|^2\bigr)^\alpha |f(z)|^p \, dA(z) < \infty,$$

where $dA(z)$ is the normalized Lebesgue area measure on $\mathbb{D}$. The weighted Bergman spaces are Banach spaces under the norm $\|\cdot\|_{A^p_\alpha}$.

It is well known that A^p_α is a Banach space of analytic functions that contains H^∞ as a subspace. For the cases of an elliptic or parabolic automorphism φ, it is the case that $\rho(C_\varphi) = 1$ and the spectrum of C_φ on A^p_α is the same as for C_φ acting on $\mathcal{X}$ (see Lemma 4.2 and Theorem 4.14 of [Hyvärinen et al. 2013]). However, as was the case for the Hardy spaces, it is not the case that the spectral radius is 1 for every composition operator induced by an automorphism. In fact, if φ is hyperbolic, then

$$\rho(C_\varphi) = \max\left\{\frac{1}{\varphi'(a)^s}, \frac{1}{\varphi'(b)^s}\right\}$$

where $s = (\alpha + 2)/p$, a is the Denjoy–Wolff point and b is the other fixed point of φ (see Theorem 4.6 of [Hyvärinen et al. 2013]). In turn, the spectrum contains the annulus

$$\min\left\{\frac{1}{\varphi'(a)^s}, \frac{1}{\varphi'(b)^s}\right\} \le |z| \le \max\left\{\frac{1}{\varphi'(a)^s}, \frac{1}{\varphi'(b)^s}\right\}$$

(see Corollary 4.7 of [Hyvärinen et al. 2013]).

4B3. *Weighted Banach spaces.* For $0 < p < \infty$, the standard weighted Banach space on $\mathbb{D}$, denoted $H^\infty_p = H^\infty_p(\mathbb{D})$, is the space of analytic functions on $\mathbb{D}$ such that

$$\|f\|_{H^\infty_p} = \sup_{z \in \mathbb{D}} \bigl(1 - |z|^2\bigr)^p |f(z)| < \infty,$$

The weighted Banach spaces are, not surprisingly, Banach spaces under $\|\cdot\|_{H^\infty_p}$.

It is well known that H_p^∞ is a Banach space of analytic functions that contain H^∞ as a subspace. For the cases of an elliptic or parabolic automorphism φ, it is the case that $\rho(C_\varphi)=1$ and the spectrum of C_φ on H_p^∞ is the same as for C_φ acting on $\mathcal{X}$ (see Lemma 4.2 and Theorem 4.14 of [Hyvärinen et al. 2013]). However, as was the case for the Hardy spaces, it is not the case that the spectral radius is 1 for every composition operator induced by an automorphism. In fact, if φ is hyperbolic, then

$$\rho(C_\varphi)=\max\left\{\frac{1}{\varphi'(a)^s},\frac{1}{\varphi'(b)^s}\right\}$$

where a is the Denjoy–Wolff point and b is the other fixed point of φ. In turn, the spectrum contains the annulus

$$\min\left\{\frac{1}{\varphi'(a)^p},\frac{1}{\varphi'(b)^p}\right\}\le|z|\le\max\left\{\frac{1}{\varphi'(a)^p},\frac{1}{\varphi'(b)^p}\right\}$$

(see Theorem 4.6 and Corollary 4.7 of [Hyvärinen et al. 2013]).

4B4. *Dirichlet space.* The Dirichlet space on $\mathbb{D}$, denoted $\mathcal{D}$, is the space of analytic functions on $\mathbb{D}$ such that

$$\int_{\mathbb{D}}|f'(z)|^2\,dA(z)<\infty$$

where dA denotes the normalized Lebesgue area measure on $\mathbb{D}$. Under the norm

$$\|f\|_{\mathcal{D}}^2=|f(0)|^2+\int_{\mathbb{D}}|f'(z)|^2\,dA(z)$$

the Dirichlet space has a Hilbert space structure. Although not every analytic self-map of $\mathbb{D}$ induce bounded composition operators on $\mathcal{D}$, univalent maps, and thus the automorphisms, of $\mathbb{D}$ do.

Independently, Donaway [1999, Corollary 3.11] and Martín and Vukotić [2005, Theorem 7] showed that composition operators on $\mathcal{D}$ induced by univalent self-maps of $\mathbb{D}$, and thus the automorphisms, have spectral radius 1. However, by direct calculation, one can see that the functions in $\mathcal{F}$ are not contained in the Dirichlet space; for the case of f_s this is shown in [Pons 2010] (see p. 455). Despite $\mathcal{D}$ not satisfying all the properties of $\mathcal{X}$, the spectrum of automorphism induced composition operators on $\mathcal{D}$ are precisely the same as those on $\mathcal{X}$.

To overcome the lack of eigenfunctions, the authors in [Higdon 1997] and [Gallardo-Gutiérrez and Montes-Rodríguez 2003] used two new approaches. In [Higdon 1997], the author produces approximate eigenfunctions and in [Gallardo-Gutiérrez and Montes-Rodríguez 2003] unitary similarity is the key tool.

Remark 4.2. For all of the spaces discussed in Sections 4A and 4B (and those discussed in the next section), the spectrum of C_φ when φ is elliptic will be the same as that for C_φ acting on $\mathcal{X}$. This is due to the fact that the eigenfunctions are the monomials, which are contained in all of these spaces.

5. Open questions

We end this paper with open questions which were inspired while developing the examples and comparisons in Sections 4A and 4B.

5A. ***The little Bloch space.*** While the Bloch space contains the polynomials, they are not dense in $\mathcal{B}$. The closure of the polynomials with respect to $\|\cdot\|_{\mathcal{B}}$ is called the little Bloch space, denoted $\mathcal{B}_0 = \mathcal{B}_0(\mathbb{D})$. More formally, the little Bloch space consists of the functions $f \in \mathcal{B}$ such that

$$\lim_{|z|\to 1} \left(1-|z|^2\right)\left|f'(z)\right| = 0.$$

From Theorem 12 of [Arazy et al. 1985], bounded composition operators on $\mathcal{B}_0$ are induced exactly by functions *in* $\mathcal{B}_0$, which include the automorphisms. Donaway also proved the spectral radius of every bounded composition operator on $\mathcal{B}_0$ is 1. Thus property (ii) is satisfied by $\mathcal{B}_0$. However, the following result shows that $\mathcal{F}$ is not contained in $\mathcal{B}_0$, and thus property (i) of $\mathcal{X}$ is not satisfied.

Theorem 5.1. *The functions f_s and f_t, for $s > 0$ and $t \neq 0$, are not contained in the little Bloch space.*

Proof. Consider the function

$$f_t(z) = \exp\left(it \log \frac{1+z}{1-z}\right).$$

We show that this function is not in $\mathcal{B}_0$ for $t \in \mathbb{R} \setminus \{0\}$. Taking the derivative,

$$f_t'(z) = f_t(z)\left(it\frac{1-z}{1+z}\right)\frac{2}{(1-z)^2} = f_t(z)\frac{2it}{(1-z)(1+z)}.$$

For $t > 0$, $|f_t(z)| \geq e^{-2\pi}$ and, for $t < 0$, $|f_t(z)| \geq 1$. In either case, there is a constant $C > 0$ such that $|f_t(z)| \geq C$ for all $t \in \mathbb{R} \setminus \{0\}$ and all $z \in \mathbb{D}$. Hence

$$|f_t'(z)| \geq \frac{2C|t|}{|z-1||z+1|}.$$

To show that $f_t \notin \mathcal{B}_0$, we need to show that

$$\lim_{|z|\to 1} \left(1-|z|^2\right)|f_t'(z)| \neq 0.$$

To see this, first observe that

$$\lim_{|z|\to 1} \left(1-|z|^2\right)|f_t'(z)| \geq \lim_{|z|\to 1} \left(1-|z|^2\right)\frac{2C|t|}{|z-1||z+1|}$$

by our estimate from above. If we now take a radial path to 1, that is, we set $z = r$ and let $r \uparrow 1$, we have

$$\lim_{r \to 1^-} \left(1 - r^2\right) \frac{2C|t|}{(1-r)(1+r)} = 2C|t| > 0$$

when $t \neq 0$. Thus

$$\lim_{|z| \to 1} \left(1 - |z|^2\right) \frac{2Ct}{|z-1||z+1|} \neq 0$$

for $t \neq 0$, and hence f_t is not in $\mathcal{B}_0$.

Next consider the function

$$f_s(z) = \exp \frac{s(z+1)}{z-1}.$$

We will show that this function is not in $\mathcal{B}_0$ for $s > 0$. First observe that

$$f_s'(z) = \left(\exp \frac{s(z+1)}{z-1}\right) \frac{-2s}{(z-1)^2}$$

and thus we aim to show that

$$\lim_{|z| \to 1} \left(1 - |z|^2\right) |f_s'(z)| = \lim_{|z| \to 1} \left(1 - |z|^2\right) \left|\exp(\frac{s(z+1)}{z-1}\right| \frac{2s}{|1-z|^2} \neq 0.$$

Fix $x_0 < 0$ and consider the sequence $\{z_n\}$ defined by

$$z_n = \frac{x_0 + in + 1}{x_0 + in - 1}.$$

Since $x_0 < 0$, this sequence is contained in the unit disk and $\{z_n\} \to 1$ as $n \to \infty$. To obtain our conclusion, we show

$$\lim_{n \to \infty} \left(1 - |z_n|^2\right) \left|\exp \frac{s(z_n+1)}{z_n - 1}\right| \frac{2s}{|1-z_n|^2} \neq 0.$$

First observe that the map $\psi(z) = (z+1)/(z-1)$ is its own inverse and hence $\psi(z_n) = x_0 + in$ for each $n \in \mathbb{N}$. Thus

$$\left|\exp \frac{s(z_n+1)}{z_n-1}\right| = \left|\exp(sx_0 + isn)\right| = e^{sx_0} > 0.$$

Substituting,

$$\lim_{n \to \infty} \left(1 - |z_n|^2\right) \left|\exp \frac{s(z_n+1)}{z_n-1}\right| \frac{2s}{|1-z_n|^2} = \lim_{n \to \infty} e^{sx_0} \left(1 - |z_n|^2\right) \frac{2s}{|1-z_n|^2}.$$

Next,

$$1 - |z_n|^2 = \frac{-4x_0}{(x_0-1)^2 + n^2}$$

and

$$|1-z_n|^2 = \frac{4}{(x_0-1)^2+n^2}.$$

Thus

$$\begin{aligned}\lim_{n\to\infty} e^{sx_0}\bigl(1-|z_n|^2\bigr)\frac{2s}{|1-z_n|^2} &= \lim_{n\to\infty} e^{sx_0}\left(\frac{-4x_0}{(x_0-1)^2+n^2}\right)\left(\frac{s\bigl((x_0-1)^2+n^2\bigr)}{2}\right)\\ &= \lim_{n\to\infty}(-2sx_0)e^{sx_0} > 0\end{aligned}$$

and hence f_s is not in $\mathcal{B}_0$ for $s > 0$. □

For the little Bloch space, we leave the reader with the following question.

Question 1. For φ a parabolic or hyperbolic automorphism, what is the spectrum of C_φ on the little Bloch space?

5B. ***Analytic functions of vanishing mean oscillation.*** Like the Bloch space, the polynomials are contained in BMOA, but they are not dense in BMOA. We denote by VMOA the closure of the polynomials in $\|\cdot\|_{\mathrm{BMOA}}$. VMOA is the space of analytic functions with vanishing mean oscillation on $\partial\mathbb{D}$, formally defined as the functions $f \in$ BMOA such that

$$\lim_{|a|\to 1}\bigl\|f\circ\tau_a - f(a)\bigr\|_{H^2} = 0.$$

By Corollary 4.2 of [Laitila 2009], C_φ is bounded on VMOA if and only if $\varphi \in$ VMOA. So every automorphism induces a bounded composition operator on VMOA. By the same argument as in Section 4A3, the spectral radius of C_φ induced by a disk automorphism is 1. Thus property (ii) of $\mathcal{X}$ is satisfied. Since VMOA is a subspace of the little Bloch space (see [Gallardo-Gutiérrez et al. 2013]), it follows that VMOA does not satisfy property (i), a corollary of Theorem 5.1.

Corollary 5.2. *The functions f_s and f_t, for $s > 0$ and $t \neq 0$, are not contained in* VMOA.

For VMOA, we leave the reader with the following question.

Question 2. For φ parabolic or hyperbolic automorphism, what is the spectrum of C_φ on VMOA?

Acknowledgements

The authors would like to thank the referees for their careful review of the manuscript and their many suggestions, especially to generalize the results to a more general family of Banach spaces. The work of the second author was conducted while an undergraduate student at the University of Wisconsin-La Crosse and funded by the College of Science and Health Dean's Distinguished Fellowship.

References

[Arazy et al. 1985] J. Arazy, S. D. Fisher, and J. Peetre, "Möbius invariant function spaces", *J. Reine Angew. Math.* **363** (1985), 110–145. MR 814017 Zbl 0566.30042

[Cowen and MacCluer 1995] C. C. Cowen and B. D. MacCluer, *Composition operators on spaces of analytic functions*, Studies in Advanced Mathematics **20**, CRC Press, Boca Raton, FL, 1995. MR 1397026 Zbl 0873.47017

[Donaway 1999] R. L. Donaway, *Norm and essential norm estimates of composition operators on Besov-type spaces*, ProQuest LLC, Ann Arbor, MI, 1999. Thesis (Ph.D.)–University of Virginia. MR 2698630

[Duren 1970] P. L. Duren, *Theory of H^p spaces*, Pure and Applied Mathematics **Vol. 38**, Academic Press, New York, 1970. MR 0268655 Zbl 0215.20203

[Gallardo-Gutiérrez and Montes-Rodríguez 2003] E. A. Gallardo-Gutiérrez and A. Montes-Rodríguez, "Adjoints of linear fractional composition operators on the Dirichlet space", *Math. Ann.* **327**:1 (2003), 117–134. MR 2005124 Zbl 1048.47016

[Gallardo-Gutiérrez et al. 2013] E. A. Gallardo-Gutiérrez, M. J. González, F. Pérez-González, C. Pommerenke, and J. Rättyä, "Locally univalent functions, VMOA and the Dirichlet space", *Proc. Lond. Math. Soc.* (3) **106**:3 (2013), 565–588. MR 3048550 Zbl 1276.30026

[Higdon 1997] W. M. Higdon, *Composition operators of the Dirichlet space*, Ph.D. thesis, Michigan State University, 1997.

[Hyvärinen et al. 2013] O. Hyvärinen, M. Lindström, I. Nieminen, and E. Saukko, "Spectra of weighted composition operators with automorphic symbols", *J. Funct. Anal.* **265**:8 (2013), 1749–1777. MR 3079234 Zbl 1325.47054

[Laitila 2009] J. Laitila, "Weighted composition operators on BMOA", *Comput. Methods Funct. Theory* **9**:1 (2009), 27–46. MR 2478261 Zbl 1163.47018

[MacCluer 2009] B. D. MacCluer, *Elementary functional analysis*, Graduate Texts in Mathematics **253**, Springer, New York, 2009. MR 2462971 Zbl 1170.46002

[Martín and Vukotić 2005] M. J. Martín and D. Vukotić, "Norms and spectral radii of composition operators acting on the Dirichlet space", *J. Math. Anal. Appl.* **304**:1 (2005), 22–32. MR 2124646 Zbl 1071.47028

[Nordgren 1968] E. A. Nordgren, "Composition operators", *Canad. J. Math.* **20** (1968), 442–449. MR 0223914 Zbl 0161.34703

[Pons 2007] M. A. Pons, *Composition Operators on Besov and Dirichlet Type Spaces*, Ph.D. thesis, University of Virginia, 2007.

[Pons 2010] M. A. Pons, "The spectrum of a composition operator and Calderón's complex interpolation", pp. 451–467 in *Topics in operator theory. Volume 1. Operators, matrices and analytic functions*, Oper. Theory Adv. Appl. **202**, Birkhäuser, Basel, 2010. MR 2723292 Zbl 1217.47050

Received: 2015-06-24 Revised: 2015-08-26 Accepted: 2015-09-07

rallen@uwlax.edu *Department of Mathematics and Statistics, University of Wisconsin-La Crosse, La Crosse, WI 54601, United States*

thmle@ucdavis.edu *Department of Computer Science, University of California, Davis, Davis, CA 95616, United States*

mapons@noctrl.edu *Department of Mathematics, North Central College, Naperville, IL 60540, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.831

On nonabelian representations of twist knots

James C. Dean and Anh T. Tran

(Communicated by Jim Hoste)

We study representations of the knot groups of twist knots into $SL_2(\mathbb{C})$. The set of nonabelian $SL_2(\mathbb{C})$ representations of a twist knot K is described as the zero set in $\mathbb{C}\times\mathbb{C}$ of a polynomial $P_K(x, y) = Q_K(y)+x^2 R_K(y) \in \mathbb{Z}[x, y]$, where x is the trace of a meridian. We prove some properties of $P_K(x, y)$. In particular, we prove that $P_K(2, y) \in \mathbb{Z}[y]$ is irreducible over $\mathbb{Q}$. As a consequence, we obtain an alternative proof of a result of Hoste and Shanahan that the degree of the trace field is precisely two less than the minimal crossing number of a twist knot.

1. Introduction

Let $J(k, l)$ be the two-bridge knot/link in Figure 1, where $k, l \neq 0$ denote the numbers of half-twists in the boxes. Positive (resp. negative) numbers correspond to right-handed (resp. left-handed) twists. Note that $J(k, l)$ is a knot if and only if kl is even. The knots $J(2, 2n)$, where $n \neq 0$, are known as twist knots. Moreover, $J(2, 2)$ is the trefoil knot and $J(2, -2)$ is the figure-eight knot. For more information about $J(k, l)$, see [Hoste and Shanahan 2004].

We study representations of the knot groups of twist knots into $SL_2(\mathbb{C})$, where $SL_2(\mathbb{C})$ denotes the set of all 2×2 matrices with determinant 1. From now on we fix a twist knot $J(2, 2n)$. By [Hoste and Shanahan 2001] the knot group of $J(2, 2n)$ has a presentation $\pi_1(J(2, 2n)) = \langle c, d \mid cu = ud\rangle$, where c, d are meridians and $u = (cd^{-1}c^{-1}d)^n$. This presentation is closely related to the standard presentation of the knot group of a two-bridge knot. Note that $J(2, 2n)$ is the twist knot K_{2n} in [Hoste and Shanahan 2001]. In this note we will follow [Tran 2015b, Lemma 1.1] and use a different presentation,

$$\pi_1(J(2, 2n)) = \langle a, b \mid aw = wb\rangle,$$

where a, b are meridians and $w = (ab^{-1})^{-n}a(ab^{-1})^n$. This presentation has been shown to be useful for studying invariants of twist knots; see [Nagasato and Tran 2013; Tran 2013a; 2015a; 2015b].

MSC2010: primary 57N10; secondary 57M25.
Keywords: Chebychev polynomial, nonabelian representation, parabolic representation, trace field, twist knot.

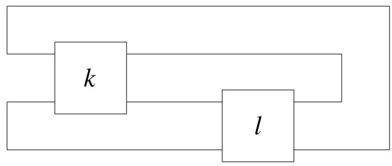


Figure 1. The two-bridge knot/link $J(k, l)$.

A representation $\rho : \pi_1(J(2, 2n)) \to \mathrm{SL}_2(\mathbb{C})$ is called nonabelian if the image of ρ is a nonabelian subgroup of $\mathrm{SL}_2(\mathbb{C})$. Suppose $\rho : \pi_1(J(2, 2n)) \to \mathrm{SL}_2(\mathbb{C})$ is a nonabelian representation. Up to conjugation, we may assume that

$$\rho(a) = \begin{bmatrix} s & 1 \\ 0 & s^{-1} \end{bmatrix} \quad \text{and} \quad \rho(b) = \begin{bmatrix} s & 0 \\ 2-y & s^{-1} \end{bmatrix},$$

where $s \neq 0$ and $y \neq 2$ satisfy a polynomial equation $P_n(s, y) = 0$. The polynomial P_n can be chosen so that $P_n(s, y) = P_n(s^{-1}, y)$, and hence it can be considered as a polynomial in the variables $x := s + s^{-1}$ and y. Note that $x = \operatorname{tr} \rho(a) = \operatorname{tr} \rho(b)$ and $y = \operatorname{tr} \rho(ab^{-1})$. An explicit formula for $P_n(x, y)$ will be derived in Section 2 and it is given by

$$P_n(x, y) = 1 - (y + 2 - x^2) S_{n-1}(y)(S_{n-1}(y) - S_{n-2}(y)),$$

where the $S_k(z)$ are the Chebychev polynomials of the second kind defined by $S_0(z) = 1$, $S_1(z) = z$ and $S_k(z) = z S_{k-1}(z) - S_{k-2}(z)$ for all integers k. Note that $P_n(x, y)$ is different from the Riley polynomial [1984] of the two-bridge knot $J(2, 2n)$; see, e.g., [Nagasato and Tran 2013]. Moreover, $P_n(2, y)$ is also different from the polynomial $\Phi_{-n}(y)$ studied in [Hoste and Shanahan 2001].

In this note we prove the following two properties of $P_n(x, y)$.

Theorem 1. *Suppose $x_0^2 \in \mathbb{R}$ such that $4 - 1/|n| < x_0^2 \leq 4$. Then the polynomial $P_n(x_0, y)$ has no real roots y if $n < 0$, and has exactly one real root y if $n > 0$.*

Theorem 2. *The polynomial $P_n(2, y) \in \mathbb{Z}[y]$ is irreducible over $\mathbb{Q}$.*

A nonabelian representation $\rho : \pi_1(J(2, 2n)) \to \mathrm{SL}_2(\mathbb{C})$ is called parabolic if the trace of a meridian is equal to 2. The zero set in $\mathbb{C}$ of the polynomial $P_n(2, y)$ describes the set of all parabolic representations of the knot group of $J(2, 2n)$ into $\mathrm{SL}_2(\mathbb{C})$. Theorem 1 is related to the problem of determining the existence of real parabolic representations in the study of the left-orderability of the fundamental groups of cyclic branched covers of two-bridge knots; see [Hu 2015; Tran 2015a].

As in the proof of [Hoste and Shanahan 2001, Theorem 1], Theorem 2 gives an alternative proof of a result of Hoste and Shanahan that the degree of the trace field is precisely two less than the minimal crossing number of a twist knot. Indeed, by definition the trace field of a hyperbolic knot K is the extension field

$\mathbb{Q}(\operatorname{tr}\rho_0(g) : g \in \pi_1(K))$, where $\rho_0 : \pi_1(K) \to \mathrm{SL}_2(\mathbb{C})$ is a discrete faithful representation. The representation ρ_0 is a parabolic representation. Since $P_n(2, y)$ is irreducible over $\mathbb{Q}$, the trace field of the twist knot $J(2, 2n)$ is $\mathbb{Q}(y_0)$, where y_0 is a certain complex root of $P_n(2, y)$ corresponding to the presentation ρ_0. Consequently, the degree of $P_n(2, y)$ gives the degree of the trace field. The conclusion follows, since the minimal crossing number of $J(2, 2n)$ is $2n+1$ if $n > 0$ and is $2-2n$ if $n < 0$.

The rest of this note is devoted to the proofs of Theorems 1 and 2.

2. Proofs of Theorems 1 and 2

In this section we first recall some properties of the Chebychev polynomials $S_k(z)$. We then compute the polynomial $P_n(x, y)$. Finally, we prove Theorems 1 and 2.

Chebychev polynomials. Recall that the $S_k(z)$ are the Chebychev polynomials defined by $S_0(z) = 1,\ S_1(z) = z$ and $S_k(z) = zS_{k-1}(z) - S_{k-2}(z)$ for all integers k. Note that $S_k(2) = k + 1$ and $S_k(-2) = (-1)^k(k + 1)$. Moreover, if $z = t + t^{-1}$, where $t \neq \pm 1$, then

$$S_k(z) = \frac{t^{k+1} - t^{-(k+1)}}{t - t^{-1}}.$$

It is easy to see that $S_{-k}(z) = -S_{k-2}(z)$ for all integers k.

The following lemma is elementary; see, e.g., [Tran 2013b, Lemma 1.4].

Lemma 2.1. *One has*

$$S_k^2(z) - zS_k(z)S_{k-1}(z) + S_{k-1}^2(z) = 1$$

for all integers k.

Lemma 2.2. *For all $k \geq 1$ one has*

$$S_k(z) = \prod_{j=1}^{k}\left(z - 2\cos\frac{j\pi}{k+1}\right),$$

$$S_k(z) - S_{k-1}(z) = \prod_{j=1}^{k}\left(z - 2\cos\frac{(2j-1)\pi}{2k+1}\right).$$

Proof. We prove the second formula. The first one can be proved similarly.

Since $S_k(z) - S_{k-1}(z)$ is a polynomial of degree k, it suffices to show that its roots are

$$2\cos\frac{(2j-1)\pi}{2k+1},$$

where $1 \leq j \leq k$. Let

$$\theta_j = \frac{(2j-1)\pi}{2k+1}.$$

Then $e^{i(2k+1)\theta_j} = -1$. Hence, if $z = 2\cos\theta_j = e^{i\theta_j} + e^{-i\theta_j}$ then we have

$$S_k(z) = \frac{e^{i(k+1)\theta_j} - e^{-i(k+1)\theta_j}}{e^{i\theta_j} - e^{-i\theta_j}} = \frac{-e^{-ik\theta_j} + e^{ik\theta_j}}{e^{i\theta_j} - e^{-i\theta_j}} = S_{k-1}(z).$$

This means that $z = 2\cos\theta_j$ is a root of $S_k(z) - S_{k-1}(z)$. □

Lemma 2.3. *Suppose $z \in \mathbb{R}$ such that $-2 \le z \le 2$. Then*

$$|S_{k-1}(z)| \le |k|$$

for all integers k.

Proof. See [Tran 2015a, Lemma 2.6]. □

Lemma 2.4. *Suppose $M \in \mathrm{SL}_2(\mathbb{C})$. Then*

$$M^k = S_{k-1}(z)M - S_{k-2}(z)I$$

for all integers k, where I is the 2×2 identity matrix and $z := \operatorname{tr} M$.

Proof. Since $\det M = 1$, by the Cayley–Hamilton theorem we have $M^2 - zM + I = 0$. This implies that $M^k - zM^{k-1} + M^{k-2} = 0$ for all integers k. Then, by induction on k we have $M^k = S_{k-1}(z)M - S_{k-2}(z)I$ for all $k \ge 0$.

For $k < 0$, since $\operatorname{tr} M^{-1} = \operatorname{tr} M = z$ we have

$$\begin{aligned} M^k = (M^{-1})^{-k} &= S_{-k-1}(z)M^{-1} - S_{-k-2}(z)I \\ &= -S_{k-1}(z)(zI - M) + S_k(z)I. \end{aligned}$$

The lemma follows, since $zS_{k-1}(z) - S_k(z) = S_{k-2}(z)$. □

The polynomial P_n. Recall that the knot group of $J(2, 2n)$ has the presentation

$$\pi_1(J(2, 2n)) = \langle a, b \mid aw = wb \rangle,$$

where a, b are meridians and $w = (ab^{-1})^{-n} a (ab^{-1})^n$. See [Tran 2015b, Lemma 1.1].

Suppose $\rho : \pi_1(J(2, 2n)) \to \mathrm{SL}_2(\mathbb{C})$ is a nonabelian representation. Up to conjugation, we may assume that

$$\rho(a) = \begin{bmatrix} s & 1 \\ 0 & s^{-1} \end{bmatrix} \quad \text{and} \quad \rho(b) = \begin{bmatrix} s & 0 \\ 2-y & s^{-1} \end{bmatrix},$$

where $s \ne 0$ and $y \ne 2$ satisfy a polynomial equation $P_n(s, y) = 0$. We now compute the polynomial P_n from the matrix equation $\rho(aw) = \rho(wb)$.

Since

$$\rho(ab^{-1}) = \begin{bmatrix} y-1 & s \\ s^{-1}(y-2) & 1 \end{bmatrix},$$

by Lemma 2.4 we have

$$\begin{aligned}\rho((ab^{-1})^n) &= S_{n-1}(y)\rho(ab^{-1}) - S_{n-2}(y)I \\ &= \begin{bmatrix} (y-1)S_{n-1}(y) - S_{n-2}(y) & sS_{n-1}(y) \\ s^{-1}(y-2)S_{n-1}(y) & S_{n-1}(y) - S_{n-2}(y)\end{bmatrix}.\end{aligned}$$

Hence, by a direct (but lengthy) calculation we have

$$\begin{aligned}\rho(aw) - \rho(wb) &= \rho\big(a(ab^{-1})^{-n}a(ab^{-1})^n\big) - \rho\big((ab^{-1})^{-n}a(ab^{-1})^n b\big) \\ &= \begin{bmatrix} (y-2)P_n(s,y) & sP_n(s,y) \\ -s^{-1}(y-2)P_n(s,y) & 0\end{bmatrix},\end{aligned}$$

where

$$P_n(s,y) = (s^2+s^{-2}+1-y)S^2_{n-1}(y) - (s^2+s^{-2})S_{n-1}(y)S_{n-2}(y) + S^2_{n-2}(y).$$

By Lemma 2.1 we have $S^2_{n-1}(y) - yS_{n-1}(y)S_{n-2}(y) + S^2_{n-2}(y) = 1$. Hence

$$P_n(s,y) = 1 - (y - s^2 - s^{-2})S_{n-1}(y)(S_{n-1}(y) - S_{n-2}(y)).$$

Since $P_n(s,y) = P_n(s^{-1},y)$, from now on we consider P_n as a polynomial in the variables $x = s + s^{-1}$ and y. With these new variables we have

$$P_n(x,y) = 1 - (y + 2 - x^2)S_{n-1}(y)(S_{n-1}(y) - S_{n-2}(y)).$$

Proof of Theorem 1. We first prove the following lemma.

Lemma 2.5. *Suppose $x_0^2 \in \mathbb{R}$ such that $4 - 1/|n| < x_0^2 \le 4$. If $y \in \mathbb{R}$ satisfies $P_n(x_0, y) = 0$, then $y > 2$.*

Proof. Since $P_n(x_0,y) = 0$, we have $S_{n-1}(y)(S_{n-1}(y) - S_{n-2}(y)) = (y+2-x_0^2)^{-1}$. Hence

$$\begin{aligned}\big((y+2-x_0^2)S_{n-1}(y)\big)^{-2} &= (S_{n-1}(y) - S_{n-2}(y))^2 \\ &= 1 + (y-2)S_{n-1}(y)S_{n-2}(y) \\ &= 1 + (y-2)\big(S^2_{n-1}(y) - (y+2-x_0^2)^{-1}\big),\end{aligned}$$

which implies that

$$1 = (y+2-x_0^2)(4-x_0^2)S^2_{n-1}(y) + (y-2)(y+2-x_0^2)^2 S^4_{n-1}(y).$$

Assume $y \le 2$. Then it follows from the above equation that

$$1 \le (y+2-x_0^2)(4-x_0^2)S^2_{n-1}(y). \tag{2-1}$$

In particular, $y > x_0^2 - 2 > -2$. Since $-2 < y \le 2$, by Lemma 2.3 we have $S^2_{n-1}(y) \le n^2$. Hence

$$(y+2-x_0^2)(4-x_0^2)S^2_{n-1}(y) \le (4-x_0^2)^2 n^2 < 1.$$

This contradicts (2-1). □

We now complete the proof of Theorem 1. Suppose $x_0^2 \in \mathbb{R}$ and $4-1/|n| < x_0^2 \le 4$. By Lemma 2.5, it suffices to consider $P_n(x_0, y)$, where y is a real number greater than 2. The equation $P(x_0, y) = 0$ is equivalent to

$$x_0^2 - 4 = y - 2 - \frac{1}{S_{n-1}(y)(S_{n-1}(y) - S_{n-2}(y))}. \tag{2-2}$$

Denote by $f_n(y)$ the right-hand side of (2-2), where $y > 2$. We now use the factorizations of $S_{n-1}(y)$ and $S_{n-1}(y) - S_{n-2}(y)$ in Lemma 2.2.

If $n = -1$ then

$$f_n(y) = y - 2 + \frac{1}{y-1} > 0 \ge x_0^2 - 4.$$

Hence $f_n(y) = x_0^2 - 4$ has no solutions on $(2, \infty)$.

If $n < -1$ then, by letting $m = -n > 1$, we have

$$\begin{aligned} f_n(y) &= y - 2 + \frac{1}{S_{m-1}(y)(S_m(y) - S_{m-1}(y))} \\ &= y - 2 + \frac{1}{\prod_{k=1}^{m-1}\left(y - 2\cos\frac{k\pi}{m}\right)\prod_{l=1}^{m}\left(y - 2\cos\frac{(2l-1)\pi}{2m+1}\right)} > 0 \ge x_0^2 - 4. \end{aligned}$$

Hence $f_n(y) = x_0^2 - 4$ has no solutions on $(2, \infty)$.

If $n = 1$ then $f_n(y) = y - 3$. Since $x_0^2 > 3$, the equation $f_n(y) = x_0^2 - 4$ has a unique solution $y = x_0^2 - 1$ on $(2, \infty)$.

If $n > 1$ then we have

$$f_n(y) = y - 2 - \frac{1}{\prod_{k=1}^{n-1}\left(y - 2\cos\frac{k\pi}{n}\right)\prod_{l=1}^{n-1}\left(y - 2\cos\frac{(2l-1)\pi}{2n-1}\right)}.$$

It is easy to see that $f_n(y)$ is increasing on $(2, \infty)$. Moreover, $\lim_{y\to\infty} f_n(y) = \infty$ and $\lim_{y\to 2} f_n(y) = -1/n < x_0^2 - 4$. Hence $f_n(y) = x_0^2 - 4$ has a unique solution on $(2, \infty)$.

The proof of Theorem 1 is complete.

Proof of Theorem 2. We write $P_n(y)$ for $P_n(2, y)$. Let $y = t^2 + t^{-2}$. Then

$$\begin{aligned} P_n(y) &= (S_{n-1}(y) - S_{n-2}(y))^2 - (y-2)S_{n-1}^2(y) \\ &= \frac{(t^{2n} + t^{2-2n})^2 - t^2(t^{2n} - t^{-2n})^2}{(t^2+1)^2} \\ &= \frac{(t^{2n} + t^{2-2n} + t^{2n+1} - t^{1-2n})(t^{2n} + t^{2-2n} - t^{2n+1} + t^{1-2n})}{(t^2+1)^2}. \end{aligned}$$

Up to a factor t^k, each of $t^{2n} + t^{2-2n} + t^{2n+1} - t^{1-2n}$ and $t^{2n} + t^{2-2n} - t^{2n+1} + t^{1-2n}$ is obtained from the other by replacing t by t^{-1}. To show that $P_n(y)$ is irreducible over $\mathbb{Q}$, it suffices to show that

$$t^{4n} + t^{4n-1} + t - 1 = (t^2 + 1) Q_n(t), \tag{2-3}$$

where $Q_n(t) \in \mathbb{Z}[t]$ is irreducible over $\mathbb{Q}$.

As in the proof of [Baker and Petersen 2013, Lemma 6.8], we will use the following theorem of Ljunggren [1960]. Consider a polynomial of the form $R(t) = t^{k_1} + \varepsilon_1 t^{k_2} + \varepsilon_2 t^{k_3} + \varepsilon_3$, where $\varepsilon_j = \pm 1$ for $j = 1, 2, 3$. Then, if R has $r > 0$ roots of unity as roots then R can be decomposed into two factors, one of degree r which has these roots of unity as zeros and the other which is irreducible over $\mathbb{Q}$. Hence, to prove (2-3) it suffices to show that $\pm i$ are the only roots of unity which are roots of $t^{4n} + t^{4n-1} + t - 1$ and these occur with multiplicity 1.

Let t be a root of unity such that $t^{4n} + t^{4n-1} + t - 1 = 0$. Write $t = e^{i\theta}$, where $\theta \in \mathbb{R}$. Since $t^{2n-1} + t^{1-2n} + t^{2n} - t^{-2n} = 0$, we have

$$2\cos(2n-1)\theta + 2i \sin 2n\theta = 0,$$

which implies that both $\cos(2n-1)\theta$ and $\sin 2n\theta$ are equal to zero. There exist integers k, l such that $(2n-1)\theta = \left(k + \frac{1}{2}\right)\pi$ and $2n\theta = l\pi$. This implies that $(2k+1)/l = (2n-1)/n$. Since $(2n-1)/n$ is a reduced fraction, there exists an *odd* integer m such that $2k+1 = m(2n-1)$ and $l = mn$. Hence $\theta = \frac{1}{2} m\pi$, which implies that $t = e^{i\theta} = \pm i$. It is easy to verify that $\pm i$ are roots of $t^{4n} + t^{4n-1} + t - 1 = 0$ with multiplicity 1.

Ljunggren's theorem then completes the proof of Theorem 2.

Acknowledgements

This research was supported by the Summer 2015 Pioneer REU in the Department of Mathematical Sciences at the University of Texas at Dallas. We would like to thank the referee for carefully reading our paper and for giving helpful comments and suggestions.

References

[Baker and Petersen 2013] K. L. Baker and K. L. Petersen, "Character varieties of once-punctured torus bundles with tunnel number one", *Internat. J. Math.* **24**:6 (2013), Article ID #1350048. MR 3078072 Zbl 1275.57016

[Hoste and Shanahan 2001] J. Hoste and P. D. Shanahan, "Trace fields of twist knots", *J. Knot Theory Ramifications* **10**:4 (2001), 625–639. MR 1831680 Zbl 1003.57014

[Hoste and Shanahan 2004] J. Hoste and P. D. Shanahan, "A formula for the A-polynomial of twist knots", *J. Knot Theory Ramifications* **13**:2 (2004), 193–209. MR 2047468 Zbl 1057.57010

[Hu 2015] Y. Hu, "Left-orderability and cyclic branched coverings", *Algebr. Geom. Topol.* **15**:1 (2015), 399–413. MR 3325741 Zbl 06425407

[Ljunggren 1960] W. Ljunggren, "On the irreducibility of certain trinomials and quadrinomials", *Math. Scand.* **8** (1960), 65–70. MR 0124313 Zbl 0095.01305

[Nagasato and Tran 2013] F. Nagasato and A. T. Tran, "Some families of minimal elements for a partial ordering on prime knots", preprint, 2013. To appear in *Osaka J. Math.* arXiv 1301.0138

[Riley 1984] R. Riley, "Nonabelian representations of 2-bridge knot groups", *Quart. J. Math. Oxford Ser.* (2) **35**:138 (1984), 191–208. MR 745421 Zbl 0549.57005

[Tran 2013a] A. T. Tran, "On the twisted Alexander polynomial for representations into $SL_2(\mathbb{C})$", *J. Knot Theory Ramifications* **22**:10 (2013), Article ID #1350059. MR 3125898 Zbl 1278.57023

[Tran 2013b] A. T. Tran, "The universal character ring of some families of one-relator groups", *Algebr. Geom. Topol.* **13**:4 (2013), 2317–2333. MR 3073918 Zbl 1280.57014

[Tran 2015a] A. T. Tran, "On left-orderability and cyclic branched coverings", *J. Math. Soc. Japan* **67**:3 (2015), 1169–1178. MR 3376583 Zbl 1328.57017

[Tran 2015b] A. T. Tran, "On left-orderable fundamental groups and Dehn surgeries on knots", *J. Math. Soc. Japan* **67**:1 (2015), 319–338. MR 3304024 Zbl 06422613

Received: 2015-07-06 Revised: 2015-10-30 Accepted: 2015-11-03

jdean@udallas.edu *University of Dallas, Irving, TX 75062, United States*

att140830@utdallas.edu *Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX 75080, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.839

Envelope curves and equidistant sets

Mark Huibregtse and Adam Winchell

(Communicated by Kenneth S. Berenhaut)

Given two sets of points A and B in the plane (called the *focal sets*), the *equidistant set* (or *midset*) of A and B is the locus of points equidistant from A and B. This article studies envelope curves as realizations of focal sets. We prove two results: First, given a closed convex focal set A that lies within the convex region bounded by the graph of a concave-up function h, there is a second focal set B (an envelope curve for a suitable family of circles) such that the graph of h lies in the midset of A and B. Second, given any function $y = h(t)$ with a continuous third derivative and bounded curvature, the envelope curves A and B associated to any family of circles of sufficiently small constant radius centered on the graph of h will define a midset containing this graph.

1. Introduction

Given two sets of points A and B in the plane (called the *focal sets*), the *equidistant set* (or *midset*) of A and B is the locus of points equidistant from A and B. For this definition to make sense, we need to know how to find the distance $d_A(p)$ from a point p to a set A. Intuitively, this is the smallest distance from p to a point of A; however, since the minimum distance may not exist if A is not closed, we define

$$d_A(p) = \inf\{d(a, p) \mid a \in A\}. \tag{1}$$

If there is a point $a \in A$ such that $d_A(p) = d(p, a)$, we call a a *foot point* of p with respect to A.

We now present a few examples of equidistant sets (see Figure 1).

Example 1.1. It is well known that the locus of points equidistant from two points in the plane is the perpendicular bisector of the line segment joining the two points.

Example 1.2. It is equally well known that the locus of points equidistant from a point (the focus) and a line not containing the point (the directrix) is a parabola.

MSC2010: 51M04.
Keywords: equidistant set, envelope curve, midset.

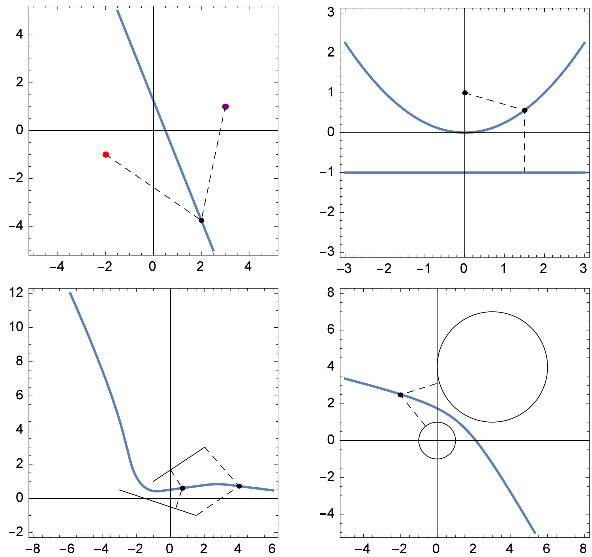


Figure 1. Clockwise from the upper left, the figures illustrate Examples 1.1, 1.2, 1.3, and 1.4.

Example 1.3. If two disjoint circles are taken as focal sets, the resulting midset is a conic section. Example 1.1 can be viewed as a special case in which the two circles have degenerated to two points, and the midset is a circle of infinite radius (the perpendicular bisector of the segment joining the two points).

Example 1.4. If the two focal sets are line segments, the midset is a curve pieced together from line segments and parabolic arcs.

The reader is encouraged to try to construct the midsets of other pairs of focal sets. It soon becomes clear that computing the exact midset of two focal sets can be a daunting problem even for relatively simple focal sets. In spite of this, certain general facts are known: For example, Ponce and Santibáñez [2014, Theorem 11, p. 28] show that under mild hypotheses on the focal sets, the equidistant sets vary continuously with the focal sets. They also suggest that the midsets associated to certain pairs of focal sets be viewed as generalized conics, by analogy with Examples 1.2 and 1.3.

In general, the results in the literature deal with the problem of describing the midset, given hypotheses on the focal sets. Our work arises from asking whether, given one focal set A and a proposed midset M, it is possible to find a second focal set B such that M is the midset (or a subset thereof) defined by A and B.

As an example, we can turn Example 1.2 on its head. Instead of being given a point (focus) and a line (directrix) and asked to find the parabolic midset they define, suppose that one is given a parabola and a point A lying in the convex region bounded by the parabola. Taking the parabola as the set M, one can ask if there

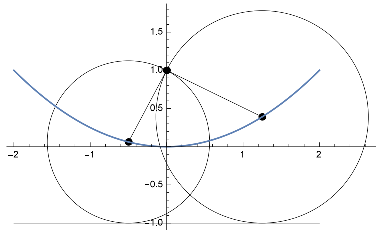


Figure 2. Another view of Example 1.2, showing how the family of circles with centers on the parabola and passing through the focus define the directrix (the line $y = -1$) as an envelope curve.

is a set B such that M is the midset of A and B. If such a set B were to exist, then for every point $m \in M$, the circle centered at m and passing through A would have to touch the boundary of B, since m must be equidistant from A and B (see Figure 2). This suggests that we can find B by constructing the envelope curves for the family of circles centered at points $m \in M$ and passing through A. In Section 4, we carry out this procedure for the more general situation in which M is the graph of a concave-up function $y = h(t)$ (that is, $h''(t) > 0$ for all t), and the point A is replaced by a closed convex set that lies "above" (in a sense to be made precise) the curve M. To set the stage, we first discuss convex sets A and the distance functions d_A they define in Section 2, and envelope curves of circles in Section 3.

In Section 5, we present our second result, which arose from asking what would happen if we started with a point A and a midset curve M that is not of uniform concavity, such as the graph of $h(t) = t^3$, and tried to find a second focal set B. We will see that the construction in Section 4 yields an envelope curve B such that the midset of A and B contains only a subset of M (not all of M). This leads to the question of how to realize the graph of $h(t) = t^3$ (or any sufficiently smooth graph M of varying concavity) as a midset. Our solution is to use the two envelope curves determined by a suitable family of circles of constant radius with centers on M: the key point is that the radius of the circles must be less than the reciprocal of the maximum curvature of M.

2. Convex sets

Recall that a set $A \subseteq \mathbb{R}^n$ is *convex* if and only if for $p, q \in A$ the segment pq lies in A. In the following discussion, we restrict attention to closed convex sets $A \subseteq \mathbb{R}^2$.

Lemma 2.1. *For any point p external to the* (*closed and convex*) *set A, there is a unique foot point $f(p)$ in A* (*in fact, in the boundary of A*).

Proof. Replacing A by the nonempty compact set

$$A' = A \cap \{p' \mid d(p, p') \le d_A(p) + 1\},$$

we see that there exists at least one point $q \in A' \subseteq A$ such that $d_A(p) = d(p, q)$, so a foot point exists. If there were two distinct foot points q_1 and q_2 for p, then the segment q_1q_2 would be a subset of A, since A is convex. From this it follows immediately that there would be a point of A lying closer to p than either foot point, which is a contradiction. Finally, it is clear that a foot point for p cannot lie in the interior of A. □

A line L is a *support line* of a plane set A if and only if L contains at least one boundary point of A and is such that the entire set A is contained in one of the two half-planes determined by L.

Lemma 2.2. *Let p be a point external to A, and $f(p) = q$ the associated foot point. Let L be the line through q that is orthogonal to the segment pq. Then L is a support line of A such that A lies in the half-plane of L that does not contain p.*

Proof. Let d be the distance from p to q. If all the points of A lie in the half-plane of L that does not contain p, we are done. If not, then there is a point q' in the (open) half-plane of L containing p. Since L is tangent at q to the circle C of radius d centered at p, the line joining q and q' intersects C at two points, and therefore the segment qq' lies partially in the interior of C. Since A is convex, qq' lies in A; this implies that there are points of A that lie closer to p than d. This is a contradiction, whence the result. □

Lemma 2.3. *The map $p \mapsto f(p) = q$, sending each point not in A to its unique foot point, is continuous.*

Proof. Let (p_n) be a sequence of points external to A that converges to p, and let (q_n) be the associated sequence of foot points. By discarding the first $N \geq 1$ terms of this sequence, we may assume that $d(p_n, p) < 1$ for all n. The triangle inequality then yields that

$$d(p_n, q) \leq d(p_n, p) + d(p, q) < d(p, q) + 1 \quad \text{for all } n.$$

We claim that $q_n \to q$. To see this, we first observe that since the distance from p_n to A is minimized at q_n,

$$d(p, q_n) \leq d(p, p_n) + d(p_n, q_n) \leq d(p, p_n) + d(p_n, q) < d(p, q) + 2.$$

In other words, the sequence (q_n) lies in the (compact) intersection of A and the disk of radius $d(p, q) + 2$ centered at p, whence the sequence has a convergent subsequence (q_{n_k}) with limit q^*. We then have

$$d(p, q^*) \leq d(p, p_{n_k}) + d(p_{n_k}, q_{n_k}) + d(q_{n_k}, q^*) \qquad \text{for all } n_k;$$

since $(p_{n_k}) \to p$ and $(q_{n_k}) \to q^*$, we obtain $d(p, q^*) \leq d(p, q)$. The uniqueness of foot points (Lemma 2.1) then yields $q^* = q$. Indeed, the same argument shows that any subsequential limit of (q_n) must equal q; it follows that (q_n) converges to q. □

Corollary 2.4. *Let $A \subseteq \mathbb{R}^2$ be closed and convex. Then the distance function d_A (restricted to points $p \notin A$) is continuous.*

Proof. Let $(p_n) \to p$ be a sequence external to A, and q_n the associated sequence of foot points. Then by the preceding lemma,

$$d_A(p_n) = d(p_n, q_n) \to d(p, q) = d_A(p),$$

which yields the result. □

The situation is even better than this; indeed, we have the following (see [Giaquinta and Modica 2012, Theorem 2.21 (Motzkin), p. 75]):

Lemma 2.5. *Let $A \subseteq \mathbb{R}^2$ be closed and convex. Then the distance function d_A (restricted to points $p \notin A$) is differentiable.* □

Before proceeding to the next section, we need one more result connected to the convexity of the set A.

Lemma 2.6. *Let $p_1 \neq p_2$ be two points external to A, and let their foot points be q_1, q_2, respectively. Then $d(q_1, q_2) \leq d(p_1, p_2)$.*

Proof. Consider the segments p_1q_1 and p_2q_2. All the points on p_iq_i $(i = 1, 2)$, except for q_i, are external to A; otherwise q_i would not be the foot point of p_i. We claim that exactly one of the following cases holds:

Case 1: One of the segments contains the other, in which case $q_1 = q_2$.

Case 2: The two segments intersect (only) at $q_1 = q_2$.

Case 3: The two segments are disjoint.

To prove the claim, we consider the various possibilities: If the two segments are disjoint, we are in Case 3. If they meet, then they either overlap or they meet in exactly one point. If they overlap, then a moment's reflection shows that we are in Case 1. If they meet in one point p^*, then it is possible that $p^* = q_1 = q_2$, which is Case 2. If we had (say) $p^* = q_1 \neq q_2$, then $p^* \in A$ would be closer to p_2 than q_2, contradicting that q_2 is the foot point of p_2. Since we are assuming $p_1 \neq p_2$, we see that if the segments meet in a single point p^* and we are not in Case 2, then p^* must be in the interior of both segments. Supposing this to be the case, we now complete the proof of the claim by deriving a contradiction: Let L denote the line perpendicular to p_1q_1 at q_1, which by Lemma 2.2 is a support line of A. Since $q_2 \in A$, we know that q_2 must lie in the half-plane of L that does not contain p_1. There are now two subcases (see Figure 3):

Subcase 1: The point p_2 lies in the same (open) half-plane of L as does p_1. In this case, we see that the angle $p_2q_1q_2$ must be obtuse, which implies that $d(p_2, q_1) < d(p_2, q_2)$, contradicting that q_2 is the foot point of p_2.

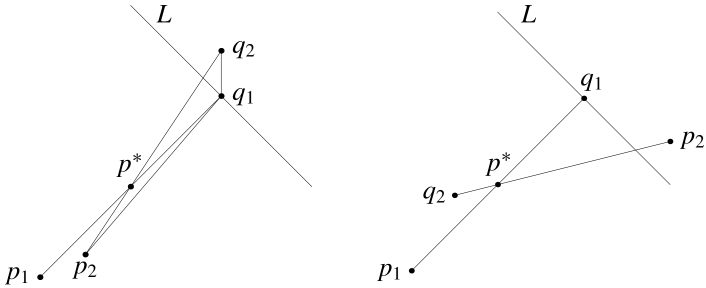


Figure 3. The two subcases of Lemma 2.6; Subcase 1 to the left and Subcase 2 to the right.

Subcase 2: The point p_2 lies in the (closed) half-plane of L that does not contain p_1. Then, since p^* is in the other half-plane of L, we see that the segment p_2p^* intersects L. From this it follows that q_2 lies in the open half-plane of L that contains p_1, which in turn implies that $q_2 \notin A$, and again we have a contradiction.

Since both cases lead to contradictions, the claim is proved.

In Cases 1 and 2, we have that $d(q_1, q_2) = 0$ and $d(p_1, p_2) > 0$, so the desired inequality is immediate. It remains to prove the inequality in Case 3. To this end, consider the segment q_1q_2 and the angles that the segments p_iq_i make with it for $i = 1, 2$ (see Figure 4). Since $q_1q_2 \subseteq A$, the support line L perpendicular to p_iq_i at q_i must be such that the segment q_1q_2 lies in the (closed) half-plane not containing p_i; this implies that the angle θ_i between p_iq_i and q_1q_2 is either right or obtuse. It is then a routine geometric exercise to show that $d(q_1, q_2) \leq d(p_1, p_2)$. $\square$

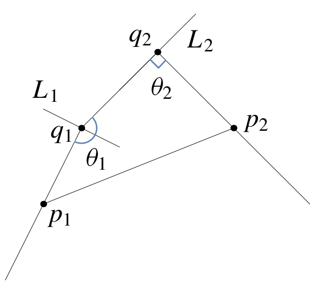


Figure 4. The angle θ_1 is drawn obtuse and the second angle θ_2 is right. One can see that if either of these angles was acute, then the segment q_1q_2 would lie on the "wrong" side of the associated support line L_i. In the case illustrated, p_1 and p_2 lie on the same side of the line through q_1 and q_2; the reader can check that the desired conclusion still holds if one or both of p_1 and p_2 are either on this line or lie on opposite sides of it.

3. Envelopes of circles

In this section we recall the definition of the envelope of a family of circles, and show how to find parametric equations of the envelope curves. We will assume that the circles have centers on the graph of a differentiable function $y = h(t)$, and radii given by a differentiable function $r(t)$. Then the family of circles is described by

$$F(x, y, t) = (x-t)^2 + (y-h(t))^2 - r^2(t) = 0; \tag{2}$$

that is, a point (x, y) lies on one of the circles if and only if there is a value of t such that (2) holds. As explained in, e.g., [Bruce and Giblin 1992, Chapter 5, pp. 99–103] or [Cox et al. 1992, Chapter 3, Section 4, pp. 139–144], the point (x, y) lies on one of the envelope curves if and only if there is a t such that both (2) and the following equation hold:

$$\frac{\partial F(x, y, t)}{\partial t} = -2(x-t) - 2(y-h(t))h'(t) - 2r(t)r'(t) = 0. \tag{3}$$

Solving (3) for x, we obtain

$$x = t - (y-h(t))h'(t) - r(t)r'(t); \tag{4}$$

substituting this into (2) yields a quadratic equation for y:

$$\bigl(-(y-h(t))h'(t) - r(t)r'(t)\bigr)^2 + (y-h(t))^2 - r^2(t) = 0. \tag{5}$$

From the quadratic formula, we obtain parametric equations for the two envelope curves ($i = 1, 2$):

$$\begin{aligned} y_i(t) &= \frac{h(t)+h(t)h'(t)^2-r(t)r'(t)h'(t)+(-1)^{i+1}\sqrt{r(t)^2(1+h'(t)^2-r'(t)^2)}}{1+h'(t)^2}, \\ x_i(t) &= t-(y_i(t)-h(t))h'(t)-r(t)r'(t). \end{aligned} \tag{6}$$

For example, when $h(t) = \sin t$ and $r(t) = 2/(2+t^2)$, the two envelope curves are shown in Figure 5. Also note that in order for the envelope curves to be defined as real curves, the expression under the radical must be nonnegative, which in turn requires that

$$1 + h'(t)^2 - r'(t)^2 \geq 0. \tag{7}$$

In other (less precise) words, the radius function cannot change too rapidly.

We conclude this section with one more lemma concerning the envelope curves of families of circles. We begin by defining the following vector-valued functions ($i = 1, 2$), viewing $\mathbb{R}^2$ as the xy-plane in $\mathbb{R}^3$:

$$\begin{aligned} \boldsymbol{v}(t) &= (1, h'(t), 0), \\ \boldsymbol{f}_i(t) &= \bigl(x_i(t)-t,\ y_i(t)-h(t),\ 0\bigr). \end{aligned} \tag{8}$$

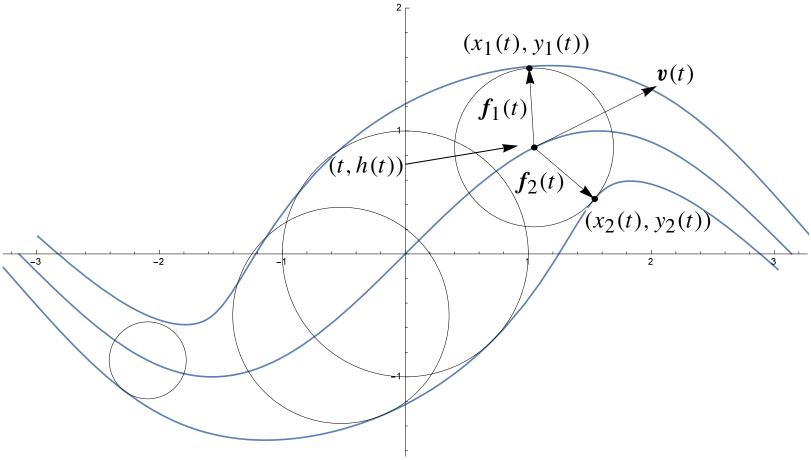


Figure 5. The envelope curves for the circles with centers on the curve $y = \sin x$ and radii given by $r(t) = 2/(2+t^2)$ are shown. The vectors $\boldsymbol{f}_i(t) = (x_i(t) - t,\, y_i(t) - h(t))$ make equal angles with the vector $\boldsymbol{v}(t) = (1, h'(t))$, as asserted by Lemma 3.1.

Lemma 3.1. *The angles between* $\boldsymbol{v}(t)$ *and the vectors* $\boldsymbol{f}_i(t)$, $i = 1, 2$, *are equal and lie on opposite sides of* $\boldsymbol{v}(t)$ *at* $(t, h(t))$ (*see* Figure 5). *Concomitantly, the cross products* $\boldsymbol{v}(t) \times \boldsymbol{f}_i(t)$ *have nonzero third components that are equal in magnitude and opposite in sign, with the sign being positive for* $i = 1$.

Proof. By direct computation, one obtains the following equations (note that the first is equivalent to (3)):

$$\begin{aligned} \boldsymbol{v}(t) \cdot \boldsymbol{f}_i(t) &= -r(t)r'(t), \\ \boldsymbol{v}(t) \times \boldsymbol{f}_i(t) &= (-1)^{i+1}\Big(0, 0, \sqrt{r(t)^2(h'(t)^2 - r'(t)^2 + 1)}\Big). \end{aligned} \tag{9}$$

Since $\boldsymbol{v}(t) \cdot \boldsymbol{f}_1(t) = \boldsymbol{v}(t) \cdot \boldsymbol{f}_2(t)$, we obtain the equality of the angles. Moreover, it is clear that the third components of the cross products $\boldsymbol{v}(t) \times \boldsymbol{f}_i(t)$, $i = 1, 2$, are equal in magnitude and opposite in sign, with the sign being positive for $i = 1$. $\square$

We will say that the envelope curve $(x_i(t), y_i(t))$ lies *above* (resp. *below*) the curve $(t, h(t))$ if and only if the sign of the third component of $\boldsymbol{v}(t) \times \boldsymbol{f}_i(t)$ is positive (resp. negative) (see Figure 5).

4. Generating a focal set B from a midset M and a focal set A

Let $y = h(t)$ be a function satisfying $h''(t) > 0$ for all t on some open interval (a, b), so that the graph of h is concave up on (a, b). (The graph of h is our intended midset M.) Let A be a closed and convex set, and f the foot-point map taking each point $p \notin A$ to its unique foot point $f(p) = q \in A$. For points $p = (t, h(t))$ on the

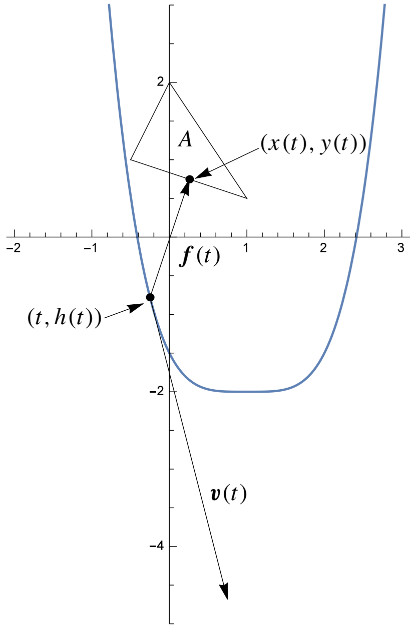


Figure 6. This picture illustrates the general situation we are considering: the closed and convex set A lies above a concave-up curve $y = h(t)$. The vectors $\boldsymbol{v}(t)$ and $\boldsymbol{f}(t)$ are shown.

graph of h, the points $f(t, h(t)) = (x(t), y(t))$ lie on a parametrized curve that we will call the *foot-point curve*.

Lemma 4.1. *The foot-point curve* $(x(t), y(t))$ *is continuous.*

Proof. Since the foot-point map f is continuous by Lemma 2.3, and the curve $(t, h(t))$ is continuous (in fact, twice differentiable) by hypothesis, we have that $(x(t), y(t))$ is the composition of continuous functions, and hence continuous. □

Let

$$\boldsymbol{f}(t) = (x(t) - t,\ y(t) - h(t),\ 0). \tag{10}$$

We assume that A lies above the graph of h in the sense that $\boldsymbol{v}(t) \times \boldsymbol{f}(t)$ is nonzero and points in the positive z-direction for all $t \in (a, b)$ (see Figure 6).

We define the radius function r as

$$r(t) = d\big((t, h(t)), (x(t), y(t))\big) = \|\boldsymbol{f}(t)\|. \tag{11}$$

Lemma 2.5 yields the following result.

Lemma 4.2. *The radius function* $r(t)$ *is* (*continuous and*) *differentiable.*

Proof. This is an immediate consequence of the fact that $r(t) = (d_A \circ h)(t)$ is a composition of differentiable functions. □

We now consider the envelope curves (6) for the family of circles with centers on the curve $y = h(t)$ and radii given by $r(t)$ for $t \in (a, b)$. Note that condition (7) holds in this case, since the rate at which $r(t)$ changes cannot possibly be greater than the speed with which the point $(t, h(t))$ moves, which is equivalent to (7). Our first main result is the following:

Proposition 4.3. *The first envelope curve* $(x_1(t), y_1(t))$ *is equal to the foot-point curve* $(x(t), y(t))$.

Proof. Assume for the moment that we know that $(x(t), y(t))$ is one of the two envelope curves. Then, since $\boldsymbol{v}(t) \times \boldsymbol{f}(t)$ has positive third component by hypothesis, Lemma 3.1 implies that $(x(t), y(t)) = (x_1(t), y_1(t))$, as desired. So we are reduced to showing that the foot-point curve $(x(t), y(t))$ satisfies conditions (2) and (3) defining the envelope curves.

It is clear that $(x(t), y(t))$ satisfies (2), since $(x(t), y(t))$ lies at distance $r(t)$ from $(t, h(t))$, by definition. We proceed to show that $(x(t), y(t))$ satisfies (3), which (as noted in the proof of Lemma 3.1) can be rearranged to read

$$r(t)r'(t) = -(1, h'(t)) \cdot (x(t)-t,\ y(t)-h(t)) = -\boldsymbol{v}(t) \cdot \boldsymbol{f}(t). \tag{12}$$

It is in fact straightforward to prove (12) at any point t at which the foot-point curve $(x(t), y(t))$ is differentiable. We simply differentiate the (differentiable, by Lemma 4.2) function $r(t)^2 = (x(t) - t)^2 + (y(t) - h(t))^2$ at t using the chain rule:

$$\begin{aligned} 2r(t)r'(t) &= 2(x(t)-t)(x'(t)-1)+2(y(t)-h(t))(y'(t)-h'(t)),\\ r(t)r'(t) &= (x(t)-t,\ y(t)-h(t))\cdot(x'(t), y'(t)) - (x(t)-t,\ y(t)-h(t))\cdot(1, h'(t)). \end{aligned}$$

We see that (3) will hold at t provided that the first term on the right-hand side of the last equation vanishes, but this follows immediately from the fact that the function

$$\boldsymbol{dd}(u) = d\big((t, h(t)), (x(u), y(u))\big)^2 = (x(u) - t)^2 + (y(u) - h(t))^2$$

has a global minimum at the point $u = t$, so that

$$\frac{d}{du}(\boldsymbol{dd})(t) = \big((x(t)-t,\ y(t)-h(t)) \cdot (x'(t), y'(t)\big) = 0. \tag{13}$$

The proof of the proposition is now complete for any case in which the foot-point curve $(x(t), y(t))$ is differentiable everywhere on its domain. (Such cases include A being a point or a disk.) Unfortunately, $(x(t), y(t))$ is not in general differentiable everywhere; however, we claim that it is always differentiable almost everywhere (that is, off of a set of measure 0). Indeed, we already know from Lemma 4.1 that $(x(t), y(t))$ is continuous, so that each component function is continuous. We will presently show that each component function has bounded variation on any closed interval $[c, d]$ in its domain, from which it follows that each component function is differentiable almost everywhere on $[c, d]$ (see, e.g., [Bressoud 2008, Theorem 7.4, p. 213]). The foregoing then implies that $(x(t), y(t))$ and the envelope curve $(x_1(t), y_1(t))$ agree almost everywhere, whence, since both of these curves are continuous, it follows that they are equal.

It remains to show that the functions $x(t)$ and $y(t)$ have bounded variation on any closed interval $[c, d]$ in the domain of h. The argument for $x(t)$ proceeds as follows; the argument for $y(t)$ is similar. By definition, we must show that there is a real number $B > 0$ such that, for every partition $P = \{t_0, t_1, \ldots, t_n\}$ of $[c, d]$,

$$\sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| \leq B.$$

However,

$$\begin{aligned}\sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| &\leq \sum_{i=1}^{n} d\big((x(t_i), y(t_i)), (x(t_{i-1}), y(t_{i-1}))\big)\\ &\leq \sum_{i=1}^{n} d\big((t_i, h(t_i)), (t_{i-1}, h(t_{i-1}))\big)\\ &\leq \int_c^d \sqrt{1 + h'(t)^2} = \text{length of curve } (t, h(t)) \text{ on } [c, d].\end{aligned}$$

The first inequality is obvious, the second follows from Lemma 2.6, and the third is due to the fact that $(t, h(t))$ is concave up, so that every polygonal approximation to its length obtained from a partition is an underestimate. □

We now want to show that the envelope curve $(x_2(t), y_2(t))$ gives us a second focal set B such that the midset of A and B contains all the points $(t, h(t))$ for $t \in (a, b)$. To do this, we must show that the distance from $(t, h(t))$ to the curve B is equal to $r(t) = d\big((t, h(t)), (x_2(t), y_2(t))\big)$, or, in other words, that the point $(x_2(t), y_2(t))$ in B is a foot point for the point $(t, h(t))$. Proposition 4.3 tells us that $(x_1(t), y_1(t)) = (x(t), y(t))$ is the foot point of $(t, h(t))$ in A, which implies that the following inequality holds for all $u, t \in (a, b)$:

$$(x_1(u) - t)^2 + (y_1(u) - h(t))^2 - \big((x_1(t) - t)^2 + (y_1(t) - h(t))^2\big) \geq 0. \tag{14}$$

Theorem 4.4. *Let B be the set of points of the envelope curve $(x_2(t), y_2(t))$. Then the point $(x_2(t), y_2(t))$ is a foot point of $(t, h(t))$ in B; consequently, $(t, h(t))$ is in the midset of A and B.*

Proof. We will show that the inequality obtained from (14) by replacing $x_1(t), y_1(t)$ by $x_2(t), y_2(t)$, respectively, holds. This will imply that the distance from $(t, h(t))$ to B is minimized at the point $(x_2(t), y_2(t))$, from which the theorem follows at

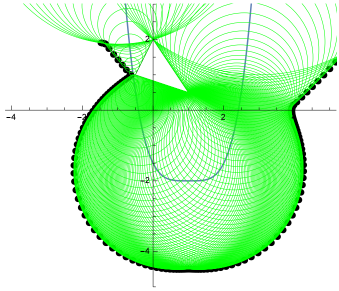


Figure 7. Another look at the example shown in Figure 6. The second envelope curve $(x_2(t), y_2(t))$ is indicated with large black dots. A subset of the circles forming the two envelope curves is shown; the result of Proposition 4.3 — that the first envelope curve $(x_1(t), y_1(t))$ is the same as the foot point curve $(x(t), y(t))$ — is also illustrated.

once. In fact, we will prove that

$$(x_2(u)-t)^2+(y_2(u)-h(t))^2-\big((x_2(t)-t)^2+(y_2(t)-h(t))^2\big)$$
$$\geq (x_1(u)-t)^2+(y_1(u)-h(t))^2-\big((x_1(t)-t)^2+(y_1(t)-h(t))^2\big)\geq 0.$$

Define $x_1(t)$, $x_2(t)$, $y_1(t)$, and $y_2(t)$ as in (6). Let

$$\boldsymbol{dd}_i(u)=(x_i(u)-t)^2+(y_i(u)-h(t))^2 \quad \text{for } i=1,2.$$

Using Mathematica to simplify $(\boldsymbol{dd}_2(u)-\boldsymbol{dd}_2(t))-(\boldsymbol{dd}_1(u)-\boldsymbol{dd}_1(t))$, we arrive at

$$-\frac{4\big(-h(t)+h(u)+(t-u)h'(u)\big)\sqrt{r(u)^2(1+h'(u)^2-r'(u)^2)}}{1+h'(u)^2}.$$

Recalling that the constraint (7) holds, the above Mathematica computation shows that the desired inequality will hold if and only if

$$-h(t)+h(u)+(t-u)h'(u)\leq 0.$$

Since this factor vanishes when $u=t$, we must show that it is nonpositive when $u\neq t$. However, this follows from our hypothesis that h is concave up ($h''>0$). Indeed, if $t<u$, then by the mean value theorem, there is a $v\in(t,u)$ such that

$$\frac{h(u)-h(t)}{u-t}=h'(v)<h'(u)\implies h'(u)(u-t)-h(u)+h(t)>0$$
$$\implies h'(u)(t-u)+h(u)-h(t)<0,$$

as desired, and a similar proof applies if $u<t$. This completes the proof of the theorem. □

Figure 7 illustrates the theorem in the case previously shown in Figure 6.

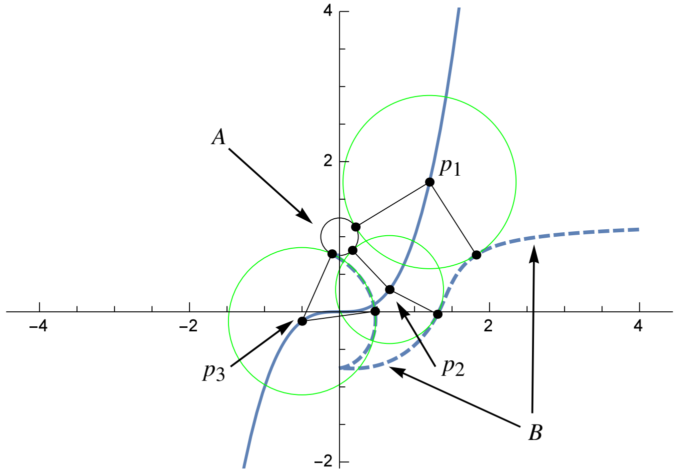


Figure 8. The first focal set A is the circle of radius $\frac{1}{4}$ centered at (0, 1). The dashed curve is the second envelope curve B for the family of circles centered on the graph of $y = t^3$ and with radii $r(t)$ equal to the distance from (t, t^3) to A. The point p_1 appears to be equidistant from A and B, but this is clearly not the case for p_2 and p_3.

5. Envelopes of curves as focal sets

The investigation in this section was undertaken in response to the question of how to realize the graph of $y = h(t) = t^3$ as a midset. The idea of the preceding section doesn't work, as Figure 8 shows. Instead, we will explore the idea of using the envelope curves A and B defined by a family of circles of constant radius centered on the graph of h as the focal sets. Figure 9 shows the envelope curves corresponding to two different radii; it becomes apparent that for this idea to work, the radius cannot be chosen too large. (Note that if $r(t) = c$ is constant, then $r'(t) = 0$ and the constraint (7) is automatically satisfied.)

We now generalize to the case of a function $y = h(t)$ with at least a continuous third derivative defined on an open interval (a, b). We no longer assume h has constant concavity, as in the preceding section. We study the envelope curves (6) defined by the family of circles of constant radius $r(t) = c$ and centered on the graph of h. It is clear that under these conditions $x_i(t)$ and $y_i(t)$ have continuous second derivatives on (a, b).

Referring to Figure 9, it appears that one constraint to the envelope curves serving as focal sets is the presence of singularities. A parametrized curve $(x(t), y(t))$ with differentiable component functions has singularities only at points where $x'(t) = 0$ and $y'(t) = 0$. We proceed to find these points on the parametrized curves

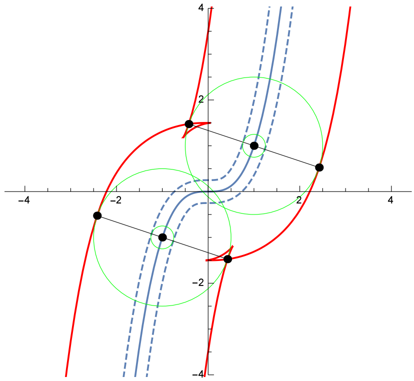


Figure 9. The envelope curves associated to the families of circles of constant radius $\frac{1}{4}$ and 1 and centered on the curve $y = t^3$. It is plausible that the dashed envelope curves have the central cubic as their midset, but this visibly cannot be the case for the solid envelope curves.

$(x_i(t), y_i(t))$; solving the system of equations $x_i'(t) = 0$ and $y_i'(t) = 0$ for c (using Mathematica), we obtain

$$c = \pm\frac{(1 + h'(t)^2)^{3/2}}{h''(t)}.$$

The keen observer will notice that the right-hand side is (up to sign) equal to the reciprocal of the curvature of the graph of h at $(t, h(t))$ (see, e.g., [Stewart 2012, Equation 11, p. 881]). To avoid singularities in the envelope curves, we should (if possible[1]) choose c such that

$$0 < c < \inf_{t\in(a,b)} \frac{(1 + h'(t)^2)^{3/2}}{|h''(t)|}. \tag{15}$$

We will refer to this upper bound for c (if it exists) as the *critical radius*; henceforth we will assume that h is such that the critical radius on (a, b) exists. Thus we have proved the following proposition.

Proposition 5.1. *If c is less than the critical radius, then the envelope curves $(x_i(t), y_i(t))$ will have no singularities.* □

For the proof of our main result, we need two more lemmas that also rely on c being less than the critical radius.

Lemma 5.2. *If c is less than the critical radius, then $x_i'(t) > 0$ for all t.*

[1]For example, the function $h(t) = \sin(1/t)$ does not have a critical radius on the interval (0, 1), since the curvature at the extreme points $t_n = 1/(n\pi + \pi/2)$ is unbounded as $n \to \infty$.

Proof. Setting $r(t) = c$ (so $r'(t) = 0$), we find that

$$x_i'(t) = \frac{(-1)^i h''(t)\sqrt{c^2(h'(t)^2+1)} + h'(t)^4 + 2h'(t)^2 + 1}{(h'(t)^2+1)^2}.$$

The quantity will be positive so long as

$$|h''(t)|\sqrt{c^2(h'(t)^2+1)} \leq (h'(t)^2+1)^2,$$

but this holds provided that

$$c \leq \frac{(h'(t)^2+1)^{3/2}}{|h''(t)|},$$

which certainly holds for all t if c is less than the critical radius. □

Lemma 5.3. *For a fixed t, let $\boldsymbol{d}_i$ denote the function*

$$\boldsymbol{d}_i(u) = d\big((t, h(t)), (x_i(u), y_i(u))\big).$$

If c is less than the critical radius, then $\boldsymbol{d}_i(t) = c$ is a local minimum value of $\boldsymbol{d}_i$; indeed, $\boldsymbol{d}_i(u) > \boldsymbol{d}_i(t) = c$ for all u in some deleted neighborhood of t.

Proof. With $r(t) = c$, we again define the function

$$\boldsymbol{dd}_i(u) := (x_i(u) - t)^2 + (y_i(u) - h(t))^2,$$

that is, $\boldsymbol{dd}_i(u) = \boldsymbol{d}_i^2(u)$. Our hypothesis on h implies that $\boldsymbol{dd}_i$ has a continuous second derivative. We find that using Mathematica to simplify $\boldsymbol{dd}_i'(t)$ returns 0 and simplifying $\boldsymbol{dd}_i''(t)$ returns

$$\frac{2\big(1 + 2h'(t)^2 + h'(t)^4 + (-1)^i\sqrt{c^2(1+h'(t)^2)}\; h''(t)\big)}{1+h'(t)^2}.$$

An argument similar to that used in the proof of the preceding lemma shows that this quantity is positive if c is less than the critical radius, whence the first assertion follows from the second derivative test. The second assertion now follows from the fact that $\boldsymbol{dd}_i''$ is continuous and positive at $u = t$, and so $\boldsymbol{dd}_i$ is concave up on a neighborhood of t. □

With these lemmas in hand we now present the main theorem of this section.

Theorem 5.4. *Given any function $h(t)$ on an interval (a, b) with a continuous third derivative and having a critical radius, the minimum distance from the point $(t, h(t))$ to either envelope curve (for the family of circles of constant radius c less than the critical radius) is equal to c, so that $(x_i(t), y_i(t))$ is a foot point for $(t, h(t))$ on the i-th envelope curve.*

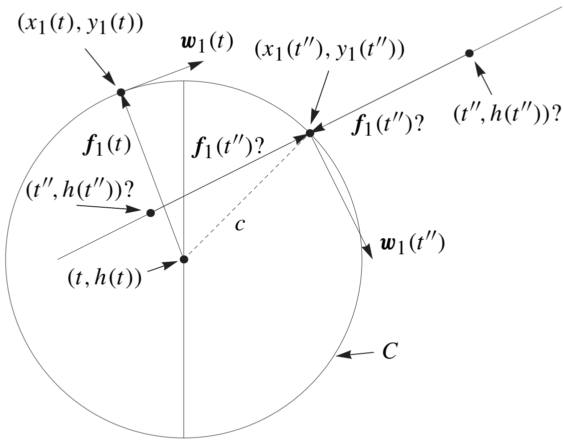


Figure 10. This figure illustrates the situation that arises in the proof of Theorem 5.4.

Proof. We begin by defining the vector-valued functions

$$\boldsymbol{w}_i(t) = (x_i'(t), y_i'(t), 0) \quad \text{for } i = 1, 2. \tag{16}$$

Machine computation yields that

$$\begin{aligned}\boldsymbol{w}_i(t) \times \boldsymbol{f}_i(t) \\ &= \left(0,\ 0,\ (-1)^{i+1} \frac{(-1)^i c^2 h''(t) + h'(t)^2 \sqrt{c^2(h'(t)^2+1)} + \sqrt{c^2(h'(t)^2+1)}}{h'(t)^2+1}\right) \\ &= \left(0,\ 0,\ (-1)^{i+1} c \frac{(-1)^i c h''(t) + (1+h'(t)^2)^{3/2}}{h'(t)^2+1}\right),\end{aligned} \tag{17}$$

where $\boldsymbol{f}_i(t)$ is defined in (8). Arguing as in the preceding lemmas, we see that whenever c is less than the critical radius, the sign $(-1)^{i+1}$ of the third component of these cross products is an invariant of the envelope curve (x_i, y_i) for $i = 1, 2$. We claim that $c = \|\boldsymbol{f}_i(t)\|$ is the global minimum distance from $(t, h(t))$ to the i-th envelope curve. We will prove this for $i = 1$, and invite the reader to check that a similar proof applies when $i = 2$.

We first observe that $\boldsymbol{w}_1(t)$ is perpendicular to $\boldsymbol{f}_1(t)$, since, by Lemma 5.3, the function $\boldsymbol{dd}_1(u)$ has a local minimum at $u = t$ (as in (13)). Hence, since $x_1'(t)$ is always positive, by Lemma 5.2, the vector $\boldsymbol{w}_1(t)$ is never parallel to the y-axis; therefore, the point $(x_1(t), y_1(t))$ lies in either the upper or lower (open) semicircle of the circle C of radius c centered at $(t, h(t))$. From (17) we learn that $\boldsymbol{w}_1(t) \times \boldsymbol{f}_1(t)$ has positive third component, whence $(x_1(t), y_1(t))$ must lie in the upper semicircle (see Figure 10). (Put more simply, $(x_1(t), y_1(t))$ is the upper envelope curve, and $(x_2(t), y_2(t))$ is the lower.)

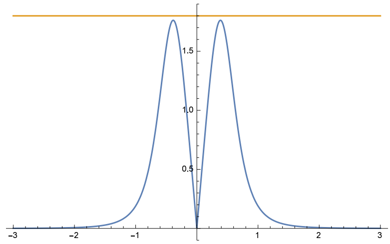


Figure 11. The graph of the curvature function $\kappa(t)$ for the function $h(t) = t^3$ is shown, along with the horizontal line $y = \frac{9}{5}$. This demonstrates that the critical radius on $(-\infty, \infty)$ exists and is slightly larger than $\frac{5}{9}$.

Arguing by contradiction, suppose that there is a point $(x_1(t'), y_1(t'))$ such that its distance from the point $(t, h(t))$ is less than c; we suppose that $t < t'$ and leave to the reader to check that a similar argument applies when $t > t'$. Since $x_1'(t)$ is always positive, we know that $(x_1(t'), y_1(t'))$ lies to the right of $(x_1(t), y_1(t))$ (indeed, this is so for any parameter value greater than t). Furthermore, by Lemma 5.3, we know that $\boldsymbol{d}_1(u) > c$ for some $u \in (t, t')$. We let

$$S = \{t^* \in (t, t') \mid \boldsymbol{d}_1(t^*) \geq c\},$$

and let $t'' = \text{lub}(S)$. By continuity, we know that $\boldsymbol{d}_1(t'') = c$, and we also know, since $\boldsymbol{d}_1(t^*) < c$ for $t^* \in (t'', t')$, that the tangent vector $\boldsymbol{w}_1(t'')$ must either be tangent to C or must enter C, as shown in Figure 10. The corresponding point $(t'', h(t''))$ must lie on the line perpendicular to the vector $\boldsymbol{w}_1(t'')$ at $(x_1(t''), y_1(t''))$, and at distance c from this point. There are consequently two possible positions for $(t'', h(t''))$, to the left or to the right of $(x_1(t''), y_1(t''))$. However, it is easy to verify that, as Figure 10 suggests, the left possibility implies that $(t'', h(t''))$ lies to the left of $(t, h(t))$ (or, more precisely, that $t'' \leq t$ holds), which contradicts $t < t''$, and that the right possibility implies that $\boldsymbol{w}_1(t'') \times \boldsymbol{f}_1(t'')$ has negative third component, which contradicts our earlier observation that this sign is always positive for (x_1, y_1). Since both possibilities lead to contradictions, we see that no point $(x_1(t'), y_1(t'))$ can lie closer than c to $(t, h(t))$, and we are done. □

Corollary 5.5. *Under the hypotheses of the theorem, the points $(t, h(t))$ all lie on the midset determined by the two envelope curves taken as focal sets.* □

Remark 5.6. For $h(t) = t^3$, the curvature function

$$\kappa(t) = \frac{|6t|}{(1 + (3t^2)^2)^{3/2}}$$

has the graph shown in Figure 11. We see that $\frac{9}{5}$ is a slight overestimate of the maximum curvature on $(-\infty, \infty)$, which implies that $\frac{5}{9}$ is a slight underestimate of the critical radius. Since $\frac{1}{4} < \frac{5}{9}$, we now have proved that the envelope curves associated to the family of circles centered on the graph of $h(t) = t^3$ and having constant radius $\frac{1}{4}$ are indeed a pair of focal sets for which the graph of h is their midset, as Figure 9 suggests.

Acknowledgment

The research leading to this article was conducted under the auspices of Skidmore College's 2014 Faculty Student Summer Research program; the authors are happy to express their gratitude for this support.

References

[Bressoud 2008] D. M. Bressoud, *A radical approach to Lebesgue's theory of integration*, Cambridge University Press, 2008. MR 2380238 Zbl 1165.00001

[Bruce and Giblin 1992] J. W. Bruce and P. J. Giblin, *Curves and singularities: a geometrical introduction to singularity theory*, 2nd ed., Cambridge University Press, 1992. MR 1206472 Zbl 0770.53002

[Cox et al. 1992] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Springer, New York, NY, 1992. MR 1189133 Zbl 0756.13017

[Giaquinta and Modica 2012] M. Giaquinta and G. Modica, *Mathematical analysis: foundations and advanced techniques for functions of several variables*, Birkhäuser, Boston, 2012. Zbl 1236.26001

[Ponce and Santibáñez 2014] M. Ponce and P. Santibáñez, "On equidistant sets and generalized conics: the old and the new", *Amer. Math. Monthly* **121**:1 (2014), 18–32. MR 3139579 Zbl 1292.51011

[Stewart 2012] J. Stewart, *Multivariable calculus*, 7th ed., Brooks/Cole, Belmont, CA, 2012.

Received: 2015-07-09 Accepted: 2015-10-20

mhuibreg@skidmore.edu *Department of Mathematics and Computer Science, Skidmore College, 815 North Broadway, Saratoga Springs, NY 12866, United States*

awinchel@skidmore.edu *Skidmore College, 815 North Broadway, Saratoga Springs, NY 12866, United States*

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.857

New examples of Brunnian theta graphs

Byoungwook Jang, Anna Kronaeur, Pratap Luitel,
Daniel Medici, Scott A. Taylor and Alexander Zupan

(Communicated by Joel Foisy)

The Kinoshita graph is the most famous example of a Brunnian theta graph, a nontrivial spatial theta graph with the property that removing any edge yields an unknot. We produce a new family of diagrams of spatial theta graphs with the property that removing any edge results in the unknot. The family is parameterized by a certain subgroup of the pure braid group on four strands. We prove that infinitely many of these diagrams give rise to distinct Brunnian theta graphs.

1. Introduction

A *spatial theta graph* is a theta graph (two vertices and three edges, each joining the two vertices) embedded in the 3-sphere S^3. There is a rich theory of spatial theta graphs and they show up naturally in knot theory. (For instance, the union of a tunnel number 1 knot with a tunnel having distinct endpoints is a spatial theta graph.) A *trivial spatial theta graph* is any spatial theta graph which is isotopic into a 2-sphere in S^3. A spatial theta graph $G \subset S^3$ has the *Brunnian property* if for each edge $e \subset G$ the knot $K_e = G \setminus e$ which is the result of removing the interior of e from G is the unknot. A spatial theta graph is *Brunnian* (or *almost unknotted* or *minimally knotted*) if it is nontrivial and has the Brunnian property.

By far the best known Brunnian theta graph is the Kinoshita graph [1958;1972]. The Kinoshita graph was generalized by Wolcott [1987] to a family of Brunnian theta graphs now called the Kinoshita–Wolcott graphs. They are pictured in Figure 1. Inspection shows that they have the Brunnian property. There are several approaches to showing that the Kinoshita graph (and perhaps all of the Kinoshita–Wolcott graphs) are nontrivial: Wolcott [1987] uses double branched covers; Litherland [1989] uses a version of the Alexander polynomial; Scharlemann [1992] and Livingston [1995] use representations of certain associated groups; McAtee et al. [2001] use quandles; Thurston [1997] showed that the Kinoshita graph is hyperbolic (i.e., the exterior supports a complete hyperbolic structure with totally geodesic boundary).

MSC2010: 57M25.

Keywords: spatial graph, almost unknotted graph, Brunnian, theta graph, theta curve, rational tangle.

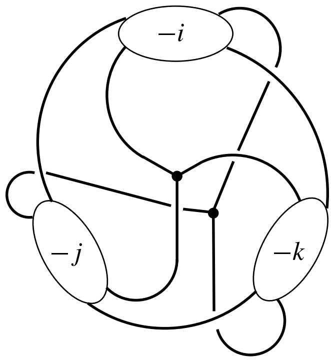


Figure 1. The Kinoshita–Wolcott graphs (figure based on [Litherland 1989, Figure 4]). The labels $-i$, $-j$, and $-k$ indicate the number of full twists in each box (with the sign of $-i$, $-j$, $-k$ indicating the direction of the twisting). If $i = j = k = 1$, the graph is the Kinoshita graph. If one of i, j or k is zero, then the graph is trivial; otherwise, it is Brunnian [Wolcott 1987, Theorem 2.1].

In this paper, we produce an infinite family of diagrams for spatial theta graphs $G(A, t_1, t_2)$ having the Brunnian property. These graphs depend on braids A lying in a certain subgroup of the pure braid group on 4 strands and on integers t_1, t_2 which represent certain twisting parameters. Our main theorem shows that infinitely many braids A give rise to Brunnian theta graphs.

Theorem 5.1 (rephrased). *For all $n \in \mathbb{Z}$, there exists a braid A_n such that for all $m \in \mathbb{Z}$, the graph $\Gamma(n, m) = G(A_n, -n, m)$ is a Brunnian theta graph. Furthermore, suppose that for a given $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, the set $S(n, m) \subset \mathbb{Z} \times \mathbb{Z}$ has the property that if $(a, b) \in S(n, m)$ then $\Gamma(a, b)$ is isotopic to $\Gamma(n, m)$ and if (a, b), $(a', b') \in S(n, m)$ are distinct, then $a + b \neq a' + b'$. Then $S(n, m)$ has at most three distinct elements. In particular, there exist infinitely many $n \in \mathbb{Z}$ such that the graphs $\Gamma(n, 0)$ are pairwise nonisotopic Brunnian theta graphs.*

2. Notation

We work in either the PL or smooth category. For a topological space X, we let $|X|$ denote the number of components of X. If $Y \subset X$ then $\eta(Y)$ is a closed regular neighborhood of Y in X and $\mathring{\eta}(Y)$ is an open regular neighborhood. More generally, $\mathring{Y}$ denotes the interior of Y.

3. Constructing new Brunnian theta graphs

There are two natural methods for constructing new Brunnian theta graphs: vertex sums and clasping.

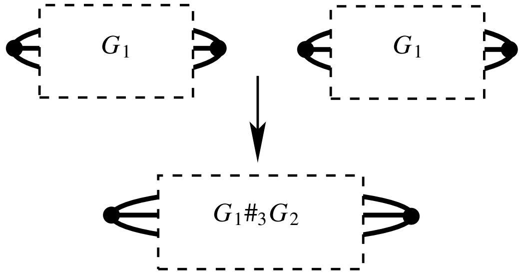


Figure 2. A schematic depiction of the vertex sum of two spatial theta graphs.

Vertex sums. Suppose that $G_1 \subset S^3$ and $G_2 \subset S^3$ are spatial theta graphs. Let $v_1 \in G_1$ and $v_2 \in G_2$ be vertices. We can construct a new spatial theta graph $G_1 \#_3 G_2 \subset S^3$ by taking the connected sum of S^3 with S^3 by removing regular open neighborhoods of v_1 and v_2 and gluing the resulting 3-balls B_1 and B_2 together by a homeomorphism $\partial B_1 \to \partial B_2$ taking the punctures $G_1 \cap \partial B_1$ to the punctures $G_2 \cap \partial B_2$. See Figure 2. The subscripted 3 represents the fact that we are performing the connected sum along a trivalent vertex and is used to distinguish the vertex sum from the connected sum of graphs occurring along edges of a graph (which, when both G_1 and G_2 are theta graphs, does not produce a theta graph).

An orientation on a spatial theta graph is a choice of one vertex to be the *source*, one vertex to be the *sink*, and a choice of a total order on the edges of the graph. If G_1 and G_2 are oriented theta graphs, we insist that the connected sum produce an oriented theta graph (so that the sink vertex of G_1 is glued to the source vertex of G_2 and so that the edges of $G_1 \#_3 G_2$ can be given an ordering which restricts to the given orderings on the edges of G_1 and G_2. Wolcott [1987] showed that the vertex sum of oriented theta graphs is independent (up to ambient isotopy of the graph) of the choice of homeomorphism $\partial B_1 \to \partial B_2$.

If G_1 and G_2 both have the Brunnian property, then $G_1 \#_3 G_2$ does as well since the connected sum of two knots is the unknot if and only if both of the original knots are unknots. If G_1 (say) is trivial, then $G_1 \#_3 G_2$ is isotopic to G_2. Similarly, if at least one of G_1 or G_2 is nontrivial then $G_1 \#_3 G_2$ is nontrivial [Wolcott 1987]. Consequently:

Theorem 3.1 (Wolcott). *If G_1 and G_2 are Brunnian theta graphs so is $G_1 \#_3 G_2$.*

We say that a spatial theta graph is *vertex-prime* if it is not the vertex sum of two other nontrivial spatial theta graphs. The Kinoshita graph is vertex prime [Calcut and Metcalf-Burton 0]. Using Thurston's hyperbolization theorem for Haken manifolds, it is possible to show that if G_1 and G_2 are theta graphs, then $G_1 \#_3 G_2$ is hyperbolic if and only if G_1 and G_2 are hyperbolic.

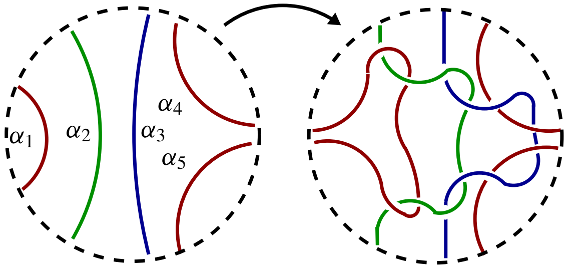


Figure 3. The clasp move in the case when all three edges of the graph are involved.

Clasping. Clasping [Simon and Wolcott 1990] is a second method for converting a Brunnian theta graph into another theta graph with the Brunnian property. Suppose that G is a spatial theta graph in S^3 which has been isotoped so that its intersection with a 3-ball $B \subset S^3$ consists of four unknotted arcs (as on the left of Figure 3), numbered α_1, α_2, α_3, α_4, and α_5. Assume that the first arc and the last two arcs belong to the same edge (the "red edge") of G and that the others belong to different, distinct edges of G (the "green edge" and the "blue edge"). We require that as we traverse the red edge, the arc α_1 is traversed between α_4 and α_5. Letting e' be the subarc of the red edge containing $\alpha_4 \cup \alpha_1 \cup \alpha_5$, we also require that there is an isotopy of e', in the complement of the rest of the graph, to an unknotted arc in B. As in Figure 3, we may then perform crossing changes to introduce a clasps between adjacent arcs. It is easily checked that this clasp move preserves the Brunnian property.

Although the clasp move creates many Brunnian theta graphs, it is not clear how to keep track of fundamental properties (such as hyperbolicity) under the clasp move. Additionally, very little is known about sequences of clasp moves relating two Brunnian theta graphs.

We can, however, use clasping to show that there exist Brunnian theta graphs which are *not* hyperbolic. Figure 4 shows an example of a Brunnian theta graph with

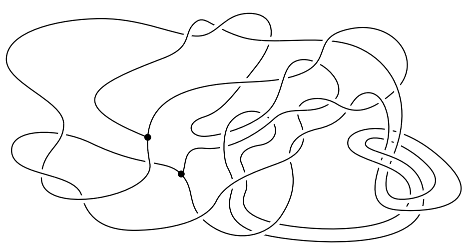


Figure 4. A toroidal Brunnian theta graph. The swallow-follow torus for the double-stranded trefoil on the right is an essential torus in the exterior of the theta graph.

an essential torus in its exterior. It was created by isotoping a Kinoshita–Wolcott graph to the position required to apply the clasping move via an isotopy which moved a point on one of the edges around a trefoil knot. A graph with an essential torus in its exterior is neither hyperbolic nor a trivial graph.

4. New examples of Brunnian theta graphs

Besides the Kinoshita graph and vertex sums of the Kinoshita graph with itself, are there other hyperbolic Brunnian theta graphs? In this section, we give a new infinite family of examples of diagrams of spatial theta curves. In the next section we will prove that infinitely many of them are also nontrivial. These examples have the property that they are of low bridge number. Forthcoming work of Taylor and Tomova will show that this implies that these graphs are vertex-prime. Furthermore, since they are low bridge number it is likely that they are hyperbolic. Section 6 concludes this paper with some questions for further research.

Let $Q = \{(x, y, z) \in \mathbb{R}^3 : -1 \le z \le 1\}$. A pure n-braid representative consists of n arcs (called strands) in Q such that the i-th strand has endpoints at $(i, 0, \pm 1)$ and for each arc projecting onto the z-axis is a strictly monotonic function. Two pure n-braid representatives are equivalent if there is an isotopy in Q from one to the other which fixes ∂Q. The set of equivalence classes is $\mathrm{PB}(n)$. Two pure n-braid representatives can be "stacked" to create another pure n-braid representative by placing one on top of the other and then scaling in the z-direction by $\frac{1}{2}$. Applying this operation to equivalence classes we obtain the group operation for $\mathrm{PB}(n)$. If σ and ρ are elements of $\mathrm{PB}(n)$, we let $\sigma\rho$ denote the braid having a representative created by stacking a representative for σ on top of a representative for ρ and then scaling in the z-direction by $\frac{1}{2}$.

Let $\phi : \mathrm{PB}(4) \to \mathrm{PB}(2)$ be the homomorphism which forgets the last two strands. For each $A \in \ker\phi$ we will construct a family $G(A, t_1, t_2)$ for $t_1, t_2 \in \mathbb{Z}$ of theta graphs with the Brunnian property. We will construct $G(A, t_1, t_2)$ by placing braids into the boxes in the template shown in Figure 5. Let $\rho : \mathrm{PB}(4) \to \mathrm{PB}(6)$ be a monomorphism which "doubles" each of the last two strands of $A \in \mathrm{PB}(4)$ (i.e., in $\rho(A)$ the fourth strand is parallel to the third and the sixth strand is parallel to the fifth). For a given $A \in \ker\phi$, we place $\rho(A)$ into the top braid box of Figure 5. The shading indicates the doubled strands. There is more than one choice for the monomorphism ρ, as the doubled strands may be allowed to twist around each other (i.e., we may vary the framing). We will always choose the homomorphism determined by the "blackboard framing" (i.e., in our diagram the doubled strands are two edges of a rectangle embedded in the plane). Into the second and fourth boxes from the top we place the braid A^{-1}. In the third box we place the element from $\mathrm{PB}(2)$ consisting of two strands with t_1 full twists. We use the convention

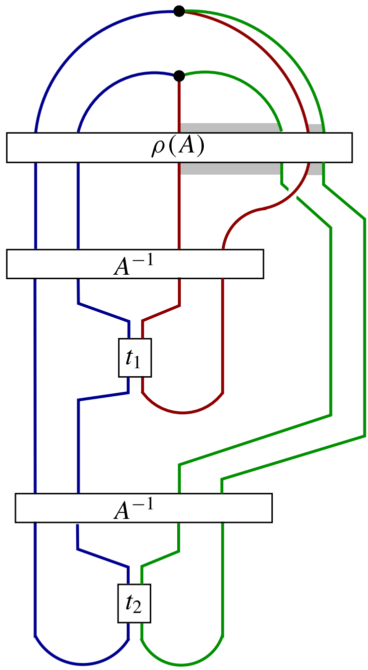


Figure 5. The template for the graph G_A.

that, giving the strands a downward orientation, if $t_1 > 0$ there are $2|t_1|$ left-handed crossings and if $t_1 < 0$ there are $2|t_1|$ right-handed twists. Into the bottom box we place t_2 full twists, using the same orientation convention as for t_1.

Considering the plane of projection in Figure 5 as the xy plane, the plane Π perpendicular to the plane of projection and cutting between the second and third boxes from the top functions as a "bridge plane" for $G(A, t_1, t_2)$. Observe that if we measure the height of a point $x \in G(A, t_1, t_2)$ by its projection onto the y-axis, each edge of $G(A, t_1, t_2)$ has a single local minimum for the height function and no other critical points in its interior. This implies that Π cuts $G(A, t_1, t_2)$ into trees with special properties. The two trees above Π have a single vertex which is not a leaf and their union is isotopic (relative to endpoints) into Π. The three trees below Π are all edges (i.e., each is a tree with two vertices and single edge) and their union can be isotoped relative to the endpoints into Π. Thus, Π is a bridge plane for $G(A, t_1, t_2)$ and $|G(A, t_1, t_2) \cap \Pi| = 6$. We might, therefore, say that $G(A, t_1, t_2)$ has "bridge number at most 3". The precise definition of bridge number for theta graphs has been a matter of dispute (see [Motohashi 2000]). The forthcoming paper of Taylor and Tomova explores bridge number for spatial graphs in detail.

Theorem 4.1. *For each $A \in \ker\phi$ and $t_1, t_2 \in \mathbb{Z}$, the graph $G(A, t_1, t_2)$ has the Brunnian property.*

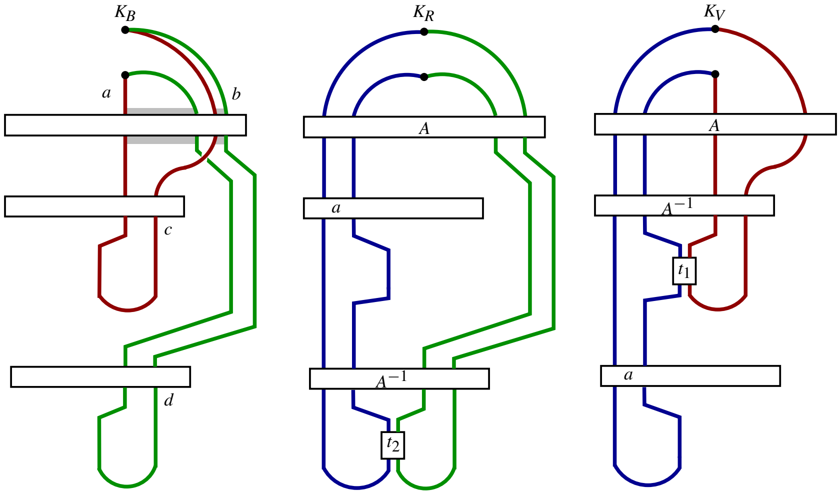


Figure 6. The constituent knots K_B, K_R, and K_V of $G(A, t_1, t_2)$.

Proof. The proof is easy and diagrammatic. Color the edges coming out of the top vertex in the diagram in Figure 5 by blue, red, and verdant from left to right. Then the edges entering into the bottom vertex are also blue, red, and verdant from left to right. In Figure 6, we have the knots K_B, K_R, and K_V obtained by removing the blue, red, and verdant edges respectively. Observe that the top braid box of K_R and K_V contains the braid A. In each of the diagrams for K_B, K_R, and K_V we have labeled certain portions with lower case letters. We now explain those regions and why each diagram can be simplified to the standard diagram for the unknot.

Consider the diagram for K_B. Since the third and fourth strands of the top braid box of $G(A, t_1, t_2)$ are parallel, we may untwist the diagram at region a and at region b. At regions c and d, we may also untwist at the minima. The end result is a diagram of a knot having a single crossing. The knot K_B must, therefore, be the unknot.

Consider the diagram for K_R. At region a we have the trivial 2-braid since $A \in \ker \phi$. The braid A in the top braid box may therefore be canceled with the braid A^{-1} in the third-from-the-top braid box. Finally, we may untwist the t_2 full twists in the final braid box to arrive at the standard diagram for the unknot.

Consider the diagram for K_V. The braids A and A^{-1} cancel, at which point we may untwist the t_1 full twists. We may also untwist at region a. Thus, K_V is also the unknot. □

Given a graph $G(A, t_1, t_2)$ we can construct other theta graphs of bridge number at most 3 with the Brunnian property by using the clasping technique in such a way that we do not introduce any additional critical points in the interior of any edge, so it is highly unlikely that the template in Figure 5 encompasses all possible theta graphs of bridge number at most 3 with the Brunnian property. On the other

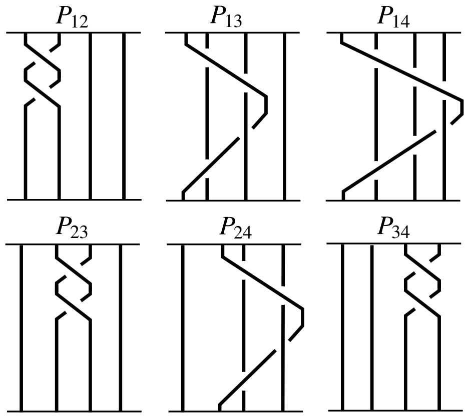


Figure 7. The generators for PB(4).

hand, there are infinitely many braids A such that $G(A, 0, 0)$ is a diagram of the trivial theta graph (see below), so the question as to what braids in $\ker\phi$ produce nontrivial theta graphs is somewhat subtle.

5. Braids producing Brunnian theta graphs

In this section we produce an infinite family of braids $A \in \ker\phi \subset \mathrm{PB}(4)$ such that there exists t_1 such that $G(A, t_1, t_2)$ is Brunnian for all t_2. To describe the braids A more precisely, we recall the standard generating set for PB(4). For $i, j \in \{1, 2, 3, 4\}$ with $i < j$, let P_{ij} denote the element of PB(4) obtained by "looping" the i-th strand around the j-th strand, as in Figure 7. Observe that P_{23}^k produces a twist box in the second and third strands with k full twists, using the sign convention from earlier.

There are nontrivial braids A for which $G(A, 0, 0)$ is trivial. For example, for every n, t_1, and t_2, the graphs $G(P_{23}^n, t_1, t_2)$ are all trivial. To show that there are infinitely many braids producing nontrivial graphs, let $A_n = P_{23}^{-n} P_{13}$ and for $m \in \mathbb{Z}$, let $\Gamma(n, m) = G(A_n, -n, m)$ (see Figure 8 for a diagram of A_2).

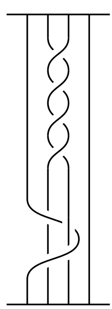


Figure 8. The braid A_2.

Theorem 5.1. *For all $n, m \in \mathbb{Z}$, the graph $\Gamma(n, m)$ is a Brunnian theta graph. Furthermore, suppose that for a given $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, the set $S(n, m) \subset \mathbb{Z} \times \mathbb{Z}$ has the properties that if $(a, b) \in S(n, m)$ then $\Gamma(a, b)$ is isotopic to $\Gamma(n, m)$ and if $(a, b), (a', b') \in S(n, m)$ are distinct, then $a + b \neq a' + b'$. Then $S(n, m)$ has at most three distinct elements. In particular, there exist infinitely many $n \in \mathbb{Z}$ such that the graphs $\Gamma(n, 0)$ are pairwise distinct Brunnian theta graphs.*

Before proving the theorem, we establish some background. A *handlebody* is the regular neighborhood of a finite graph embedded in S^3 and its *genus* is the genus of the boundary surface. We will be considering genus 2 handlebodies. A disc D properly embedded in a handlebody H whose boundary does not bound a disc in ∂H is called an *essential disc* in H. If H has genus 2 and if $D \subset H$ is an essential nonseparating disc, the space $H \setminus \mathring{\eta}(D)$ is homeomorphic to $S^1 \times D^2$. A knot isotopic to the core of that solid torus is called a *constituent knot* of H. If $G \subset S^3$ is a spatial theta graph and if $H = \eta(G)$, then a disc $D \subset H$ intersecting an edge e of G exactly once transversally and disjoint from the other edges of G is called a *meridian disc* for e. Thus, if D is a meridian disc for e, then $H \setminus \mathring{\eta}(D)$ is a regular neighborhood of K_e. Observe that if G is a theta graph and if $e \subset G$ is an edge, then any meridian disc D for e is an essential disc in the handlebody $\eta(G)$, as D does not separate $\eta(G)$.

If G and G' are spatial theta graphs such that G is isotopic to G' then the isotopy can be extended to an isotopy of the handlebody $\eta(G)$ to the handlebody $\eta(G')$. Furthermore, if the isotopy takes an edge $e \subset G$ to an edge $e' \subset G'$ then the isotopy takes any meridian disc for e to a meridian disc for e'. On the other hand, an isotopy of $\eta(G)$ to $\eta(G')$ does not necessarily correspond to an isotopy of G to G'. Instead, an isotopy of $\eta(G)$ to $\eta(G')$ corresponds to a sequence of isotopies and "edge slides" of G. An *edge slide* of an edge $e \subset G$ of a graph involves sliding one end of e across edges of G (see [Scharlemann and Thompson 1994]). As in Figure 9, an edge slide of a theta graph may convert a theta graph into a spatial graph that is not a theta graph. Conversely, any sequence of edge slides and isotopies of a graph G corresponds to an isotopy of $\eta(G)$.

Given a spatial theta graph G and an edge e, an essential nonseparating disc E in $\eta(G)$ is *along* e if it lies in a regular neighborhood of e, is not a meridian of e, if there is a meridian disc D for e such that $|D \cap E|$ (the number of components of $D \cap E$) is equal to 1. Observe that if E is along e, then $\eta(G) \setminus \mathring{\eta}$ is a solid torus

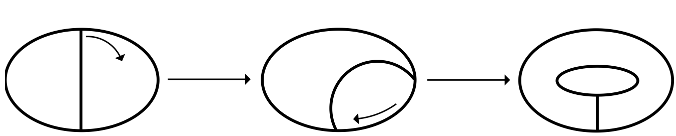


Figure 9. An edge slide converting a theta graph into a nontheta graph.

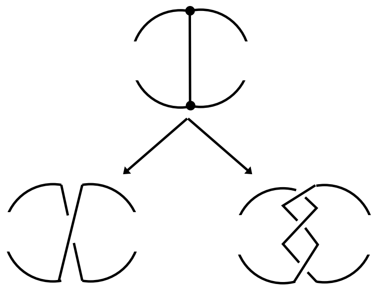


Figure 10. Two ways of unzipping an edge of a spatial theta graph. As is suggested by the picture, the θ-graph may be embedded in S^3 in some, potentially complicated, way. We do, however, require that the unzipping produce a knot and not a 2-component link.

since E is nonseparating. If E is along e, then we say that the knot which is the core of $\eta(G) \setminus \mathring{\eta}(E)$ is obtained by *unzipping* the edge e. Figure 10 shows two different ways of unzipping an edge. The proof of Lemma 5.5 will also be helpful in understanding the relationship between the definition of unzipping given above and the diagrams in Figure 10. The term "unzipping" is taken from Bar-Natan and D. Thurston (see, for example, [Thurston 2002]). It is a form of an operation also known as "attaching a band" to K_e or "distance 1 rational tangle replacement" on K_e.

Since an isotopy of a handlebody in S^3 to another handlebody takes discs in the first handlebody to discs in the second and preserves the number of intersections between discs, we have:

Lemma 5.2. *Suppose that G and G' are isotopic spatial theta graphs such that the isotopy takes an edge e of G to an edge e' of G'. If $K \subset S^3$ is a knot obtained by unzipping the edge e, then there is a knot $K' \subset S^3$ which is obtained by unzipping the edge e' such that K and K' are isotopic.*

Rational tangles. The key step in our proof of Theorem 5.1 is to show that unzipping $\Gamma(n, m)$ does not produce any knot that can be obtained by unzipping a trivial theta graph along one of its edges. Analyzing the knots we do get will show, as a by-product, that infinitely many of the $\Gamma(n, m)$ are distinct. We use rational tangles to analyze our knots.

A *rational tangle* is a pair (B, τ) where B is a 3–ball and $\tau \subset B$ is a properly embedded pair of arcs which are isotopic into ∂B relative to their endpoints. We mark the points $\partial\tau \subset \partial B$ by NW, NE, SW, and SE as in Figure 11. Two rational tangles (B, τ) and (B, τ') are *equivalent* if there is a homeomorphism of pairs $h : (B, \tau) \to (B', \tau')$ which fixes ∂B pointwise. Conway [1970] showed how to associate a rational number $r \in \mathbb{Q} \cup \{1/0\}$ to each rational tangle in such a way

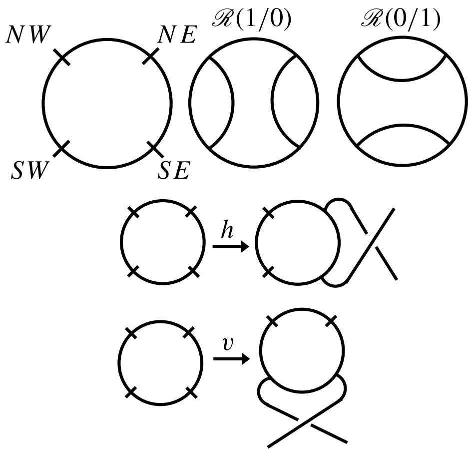


Figure 11. The basic transformations of a rational tangle.

that two rational tangles are equivalent if and only if they have the same associated rational number. We briefly explain the association, using the conventions of [Gordon 2009, Lecture 4]. Using the 3-ball with marked points as in Figure 11, we let the rational tangle $\mathscr{R}(0/1)$ consist of a pair of horizontal arcs having no crossings and we associate to it the rational number $0 = 0/1$. The rational tangle $\mathscr{R}(1/0)$, consisting of a pair of vertical arcs having no crossings, is given the rational number $1/0$ (thought of as a formal object). Let $h : B \to B$ and $v : B \to B$ be the horizontal and vertical half-twists, as shown in Figure 11. Observe that the rational tangle $v^{2k}\mathscr{R}(0/1)$ is a twist box with $-k$ full twists, using the orientation convention from earlier.

Let $a_1, a_2, \dots, a_k$ be a finite sequence of integers such that $a_2, \dots, a_k \neq 0$. Let $\mathscr{R}(a_1, \dots, a_k)$ be the rational tangle defined by

$$\mathscr{R}(a_1, a_2, \dots, a_k) = \begin{cases} h^{a_1} v^{a_2} \cdots h^{a_{k-1}} v^{a_k} \mathscr{R}(1/0) & \text{if } k \text{ is even,} \\ h^{a_1} v^{a_2} \cdots v^{a_{k-1}} h^{a_k} \mathscr{R}(0/1) & \text{if } k \text{ is odd.} \end{cases}$$

We assign the rational number

$$\frac{p}{q} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + \cfrac{1}{a_k}}}}$$

to $\mathscr{R}(a_1, a_2, \dots, a_k)$ and we define $\mathscr{R}(p/q) = \mathscr{R}(a_1, \dots, a_k)$, with p and q relatively prime.

We define the *distance* between two rational tangles $\mathscr{R}(p/q)$ and $\mathscr{R}(p'/q')$ to be $\Delta(p/q, p'/q') = |pq' - p'q|$. Observe that in the 3-ball B, there is a disc

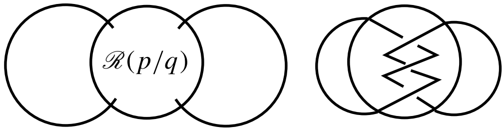


Figure 12. On the left is the denominator closure $\mathscr{D}(\mathscr{R}(p/q))$ of the rational tangle $\mathscr{R}(p/q)$. On the right, we see that the right-handed trefoil is the denominator closure of the rational tangle $\mathscr{R}(1/3)$.

$D \subset B$ such that ∂D partitions the marked points $\{NW, SW, NE, SE\}$ into pairs and which separates the strands of a given rational tangle $\mathscr{R}(p/q)$. Indeed, given a disc $D \subset B$ whose boundary partitions the marked points into pairs, there is a rational tangle $\mathscr{R}(p/q)$ (unique up to equivalence of rational tangles) such that D separates the strands of $\mathscr{R}(p/q)$. We call D a *defining disc* for $\mathscr{R}(p/q)$. If D is a defining disc for $\mathscr{R}(p/q)$ and D' is a defining disc for $\mathscr{R}(p'/q')$ such that, out of all such discs, D and D' have been isotoped to intersect minimally, then it is not difficult to show that $\Delta(p/q, p'/q') = |D \cap D'|$ (i.e., the distance between the rational tangles is equal to the minimum number of arcs of intersection between defining discs).

From a rational tangle $\mathscr{R}(p/q)$ we can create the unknot or a 2-bridge knot or link $\mathscr{K}(p/q) = \mathscr{D}(\mathscr{R}(p/q))$ by taking the so-called *denominator closure* $\mathscr{D}$ of $\mathscr{R}(p/q)$ where we attach the point NW to the point SW and the point NE to the point SE by an unknotted pair of arcs lying in the exterior of B, as in Figure 12. Thus, the right-handed trefoil is $\mathscr{K}(1/3)$ and the left-handed trefoil is $\mathscr{K}(-1/3)$.

Theorem 5.3 [Schubert 1956]. *Let p/q, $p'/q' \in \mathbb{Q} \cup \{1/0\}$ with $q, q' > 0$ and the pairs p, q and p', q' relatively prime. The knot or link $\mathscr{K}(p/q)$ is isotopic (as an unoriented knot or link) in S^3 to the knot or link $\mathscr{K}(p'/q')$ if and only if $q = q'$ and either $p \equiv p' \bmod q$ or $pp' \equiv 1 \bmod q$.*

Remark 5.4. For more on Schubert's theorem, see [Cromwell 2004, Theorem 8.7.2] or [Kauffman and Lambropoulou 2003, Theorem 3]. Since we are using the denominator closure of rational tangles our convention and the statement of Schubert's theorem differ from the usual convention and statement by exchanging numerators and denominators. See the discussion following Theorem 3 of [Kauffman and Lambropoulou 2003].

Unzipping the trivial graph. Since we want to show that each graph in a certain family of graphs is nontrivial, the following will be useful.

Lemma 5.5. *Suppose that $G \subset S^3$ is the trivial theta graph and that K is a knot obtained by unzipping an edge e of G. Then either K is the unknot or there exists $k \in \mathbb{Z}$, odd such that K is a $(2, k)$ torus knot.*

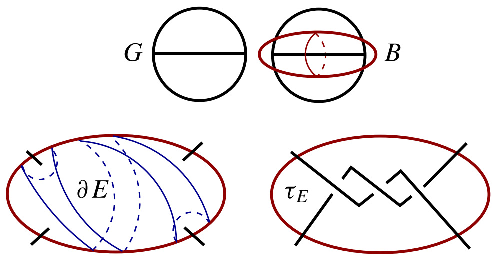


Figure 13. Upper Left: The trivial graph G. Upper Right: the ball $B = \eta(e)$. Lower Left: A disc E. Lower Right: The rational tangle $\mathscr{R}(-3)$ with defining disc E.

Proof. Let G be the trivial theta graph and let $e \subset G$ be an edge. Observe that there is an isotopy of G which interchanges any two edges. Thus, we may consider G to be the union of the unit circle in $\mathbb{R}^2$ with a horizontal diameter e, as in Figure 13. We may consider the neighborhood $\eta(e)$ of e as a 3-ball B with a vertical disc as a meridian disc for e. The graph G intersects B in four punctures, which we label NW, NE, SW, and SE as usual. Take D to be the meridian disc for e and let $E \subset B$ be a disc with boundary an essential curve in $\partial B \setminus G$, which cannot be isotoped to be disjoint from D, and for which $|D \cap E| = 1$. Observe that D is the defining disc for the rational tangle $\mathscr{R}(1/0)$. If E is the defining disc for the rational tangle $\mathscr{R}(k/\ell)$, then

$$1 = \Delta(1/0, k/\ell) = |\ell|.$$

Consequently, the rational tangle $\mathscr{R}(k/1)$ consists of k horizontal half twists. Thus the knot which is the core of $\eta(G) \setminus \mathring{\eta}(E)$ is a $(2, \pm k)$ twist knot. □

The following corollary follows immediately from Lemmas 5.2 and 5.5.

Corollary 5.6. *Suppose that $G \subset S^3$ is a trivial spatial theta graph. Then for all edges $e \subset G$ and for any knot K obtained by unzipping e there exists an odd $k \in \mathbb{Z}$ such that K is a $(2, k)$ torus knot, i.e., $\mathscr{K}(1/k)$.*

Proof of Theorem 5.1. The proof is similar in spirit to [Wolcott 1987, §3]. We do not, however, use Wolcott's theorem 3.11 as that theorem would require us to work with links, rather than with knots. Potentially, however, a clever use of Wolcott's theorem would show that a much wider class of braids A create nontrivial graphs $G(A, t_1, t_2)$. Our method, however, also allows us, using a result of Eudave-Muñoz [1992] concerning reducible surgeries on strongly invertible knots, to show that we have infinitely many distinct Brunnian theta graphs.

Let $n, m \in \mathbb{Z}$, and let $G = \Gamma(n, m)$. To prove that G is a Brunnian theta graph, by Theorem 4.1, we need only show that G is nontrivial.

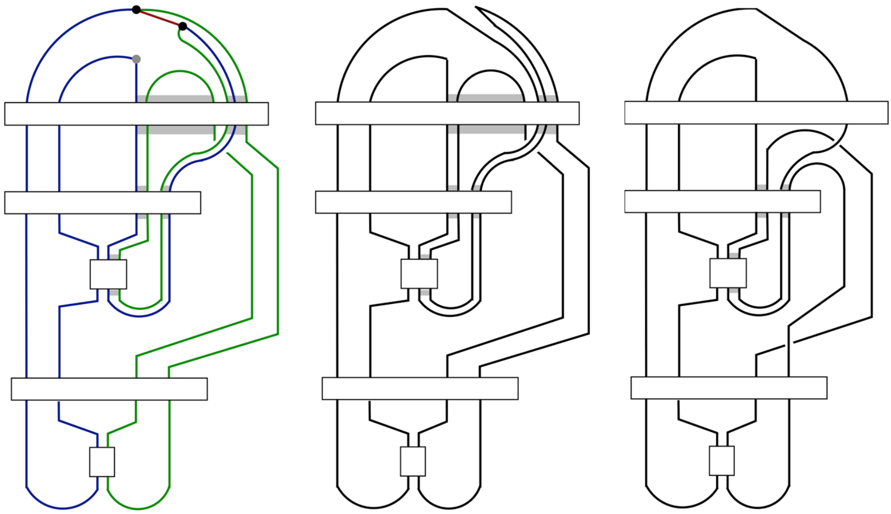


Figure 14. On the left we have isotoped the template so that R has no crossings. The shaded boxes denote doubled strands. In the middle we have unzipped along R using a particular choice of unzipping disc. On the right, we have simplified the unzipped knot by using the parallel strands from the top braid box.

Let v_+ be the upper vertex of G in Figure 5 and let v_- be the lower vertex. Recall that we color the edges of G (from left to right at each vertex) as blue, red, and verdant. Isotope G so that the endpoint of the verdant edge V adjacent to v_- is moved near v_+ by sliding it along the red edge, as on the left of Figure 14. This isotopy creates a diagram of G such that red edge has no crossings. Let K be the knot obtained by unzipping the red edge, as in the middle of Figure 14 (choosing the unzip so that no twists are inserted in the diagram along V). Using the doubled strands in the top braid box, isotope K so that it has the diagram on the right of Figure 14.

Inserting the braid A_n into the template, as specified in Figure 5, our knot K has the diagram on the top left of Figure 15. Let 3_1 be the right-handed trefoil. Now perform the isotopies indicated in Figure 15 to see that K is the connected sum of 3_1 and the knot

$$\mathscr{K}\left(-\frac{3}{6(m+n)+5}\right) = \mathscr{D}(\mathscr{R}(0, -2(m+n)-1, -1, -1, -1)).$$

Since torus knots are prime, K is not a $(2, k)$ torus knot for any $k \in \mathbb{Z}$ unless $\mathscr{K}(-3/(6(m+n)+5))$ is the trivial knot, that is $\mathscr{K}(-3/(6(m+n)+5)) = \mathscr{K}(1)$.

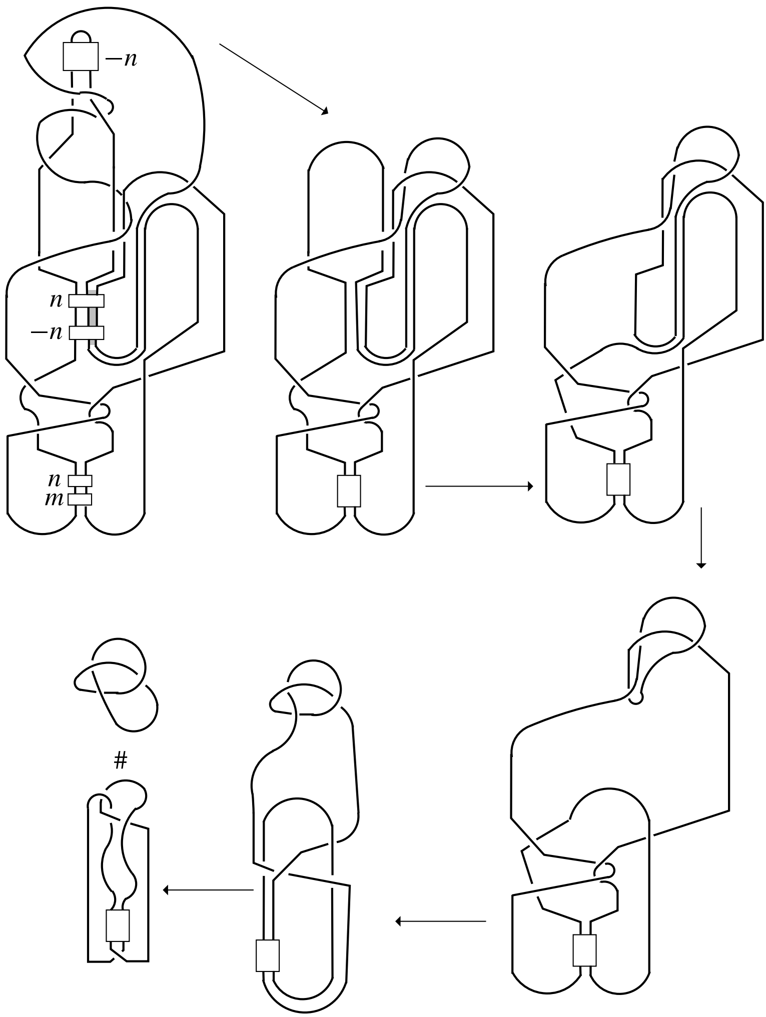


Figure 15. The isotopies showing that K is the connected sum of a right handed trefoil and $\mathscr{K}(-3/(6(m+n)+5))$. In the first step we combine the lower two twist boxes into a single twist box with $m+n$ full twists.

By Schubert's theorem, this can only happen if $6(m+n)+5=1$, an impossibility. Thus, each $\Gamma(n,m)$ is a Brunnian theta graph.

To prove the part about distinctness, we use a theorem of Eudave-Muñoz [1992] and the Montesinos trick [Montesinos 1975] (see also [Gordon 2009] for a nice explanation). We begin by showing:

Claim. *If $a+b \neq a'+b'$, then there is no isotopy from $\Gamma(a',b')$ to $\Gamma(a,b)$ which takes the red edge of $\Gamma(a',b')$ to the red edge of $\Gamma(a,b)$.*

We prove this by contradiction. Let $B \subset S^3$ be a regular neighborhood of the red edge of $\Gamma(a, b)$ and let $W = S^3 \setminus \mathring{B}$ be the complementary 3–ball. Mark the points of $\Gamma(a, b) \cap \partial B$ by NE, SE, NW, SW so that a meridian disc for the red edge of $\Gamma(a, b)$ corresponds to the rational tangle $\mathscr{R}(1/0)$ and the disc E, along which we unzip $\Gamma(a, b)$ to produce $K = 3_1 \# \mathscr{K}(-3/(6(a+b)+5))$, corresponds to the rational tangle $\mathscr{R}(0/1)$. Let $\tau = K \cap W$.

The isotopy of $\Gamma(a, b)$ to $\Gamma(a', b')$ takes B to a regular neighborhood B' of the red edge of $\Gamma(a', b')$. In B' there is a disc D' which is along the red edge of $\Gamma(a', b')$ such that unzipping $\Gamma(a', b')$ along D' produces $(3_1 \# \mathscr{K}(-3/(6(a'+b')+5)))$. Reversing the isotopy, takes D' to a disc $D \subset B$ which is along e. Let $\mathscr{R}(p/q)$ be the rational tangle corresponding to D. The knot $K' = \tau \cup \mathscr{R}(p/q))$ is isotopic to the result of unzipping $\Gamma(a', b')$ along D' and so $K' = (3_1 \# \mathscr{K}(-3/(6(a'+b')+5)))$. If the disc D is isotopic to the disc E, the rational tangles $\mathscr{R}(p/q)$ and $\mathscr{R}(0/1)$ are equivalent. In which case, K is isotopic to K'. But this implies that $a+b = a'+b'$, a contradiction. Thus, the rational tangles $\mathscr{R}(0/1)$ and $\mathscr{R}(p/q)$ are distinct (since the discs are not isotopic).

Since $\tau \cup \mathscr{R}(1/0)$ is the unknot in S^3, the double branched cover of W over τ is the exterior of a strongly invertible knot $L \subset S^3$. Since $K = \tau \cup \mathscr{R}(0/1)$ and $K' = \tau \cup \mathscr{R}(p/q)$ are composite knots, the double branched covers of S^3 over K and K' are reducible. In particular there are distinct Dehn surgeries on L producing reducible manifolds. The surgeries are distinct since $\mathscr{R}(0/1)$ is not equivalent to $\mathscr{R}(p/q)$. However this contradicts that the Cabling Conjecture holds for strongly invertible knots [Eudave-Muñoz 1992, Theorem 4]. Proving the claim. □

For a pair $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, let $S(n, m) \subset \mathbb{Z} \times \mathbb{Z}$ be a subset with the property that for all $(a, b) \in S(n, m)$, the graph $\Gamma(a, b)$ is isotopic to the graph $\Gamma(n, m)$ and which has the property that for all pairs (a, b), $(a', b') \in S(n, m)$ if $a+b = a'+b'$, then $(a, b) = (a', b')$. Observe that $(n, m) \in S(n, m)$. We will show that for all $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, the set $S(n, m)$ has at most three elements.

Suppose, for a contradiction, that there exists (n, m) such that $S(n, m)$ has at least 4 distinct elements (a_1, b_1), (a_2, b_2), (a_3, b_3), (n, m). Each isotopy between any two graphs $\{\Gamma(a_1, b_1), \Gamma(a_2, b_2), \Gamma(a_3, b_3), \Gamma(n, m)\}$ induces a permutation of the set $\{B, R, V\}$ of blue, red, and verdant edges. For each $i \in \{1, 2, 3\}$, choose an isotopy f_i from $\Gamma(n, m)$ to $\Gamma(a_i, b_i)$ and let σ_i be the induced permutation of $\{B, R, V\}$. By the claim and the definition of $S(n, m)$ no σ_i fixes R and, whenever $i \neq j$, the permutation $\sigma_i \sigma_j^{-1}$ also does not fix R. Hence, $\sigma_i \neq \sigma_j$ if $i \neq j$. In the permutations of the set $\{B, R, V\}$, there are exactly four that do not fix R and of those, two are transpositions. Thus, without loss of generality, we may assume that σ_1 is a transposition.

Suppose that σ_1 is the transposition $(B, R, V) \to (B, V, R)$. Since neither $\sigma_2 \sigma_1^{-1}$ nor $\sigma_3 \sigma_1^{-1}$ fixes R and since $\sigma_2 \neq \sigma_3$, the permutations σ_2 and σ_3 are the two permutations taking R to B. But then the composition $\sigma_2 \sigma_3^{-1}$ takes R to R, a contradiction.

The case when σ_1 is the transposition $(B, R, V) \to (R, B, V)$ similarly gives rise to a contradiction. Thus, for every $(n, m) \in \mathbb{Z} \times \mathbb{Z}$, the set $S(n, m)$ has at most three elements (including (n, m)).

Define a sequence (n_i) in $\mathbb{Z}$ recursively. Let $n_1 \in \mathbb{Z}$ and recall that, by the above, $\Gamma(n_1, 0)$ is a Brunnian theta graph. Assume we have defined $n_1, \dots, n_i$ so that the graphs $\Gamma(n_j, 0)$ for $1 \le j \le i$ are pairwise nonisotopic Brunnian theta graphs. Let $P \subset \mathbb{Z}$ be such that $n \in P$ if and only if $\Gamma(n, 0)$ is isotopic to $\Gamma(n_j, 0)$ for some $1 \le j \le i$. Since for each j with $1 \le j \le i$ there are at most 3 elements n of $\mathbb{Z}$ such that $\Gamma(n, 0)$ is isotopic to $\Gamma(n_j, 0)$, the set P is finite. Hence, we may choose $n_{i+1} \in \mathbb{Z} \setminus P$. Thus, we may construct a sequence (n_i) in $\mathbb{Z}$ so that the graphs $\Gamma(n_i, 0)$ are pairwise disjoint Brunnian theta graphs. □

6. Questions and conjectures

Using the software [Heard 2013], and a lot of patience, it is possible to compute (approximations to) hyperbolic volumes for some of the graphs $G(A, t_1, t_2)$. Our explorations suggest that "most" of the braids $A \in \mathrm{PB}(4)$ produce hyperbolic Brunnian theta graphs for all $t_1, t_2 \in \mathbb{Z}$. Indeed, the software suggests that for a "sufficiently complicated" braid $A \in \mathrm{PB}(4)$, and for fixed t_1, t_2 the volume of the exterior of $G(A^n, t_1, t_2)$ grows linearly in n. This is to be contrasted with the belief, based on the Thurston 2π theorem, that for a fixed A and t_1, the volumes of $G(A, t_1, t_2)$ will converge as $t_2 \to \infty$. Furthermore, calculations of hyperbolic volumes using Orb indicate that the graphs $\Gamma(n, m)$ of Theorem 5.1 are likely not Kinoshita–Wolcott graphs. Since the calculations of hyperbolic volume are only approximate and since we can only calculate the volumes of finitely many of the graphs, we do not have a proof of that fact.

These investigations raise the following questions:

(1) For what braids $A \in \ker \phi$ is $G(A, 0, 0)$ a Brunnian theta graph?

(2) Can Litherland's Alexander polynomial (or some other algebraic invariant) prove that there are infinitely many braids $A \in \ker \phi$ such that $G(A, t_1, t_2)$ is a Brunnian theta graph for some $t_1, t_2 \in \mathbb{Z}$?

(3) Is any one of the Brunnian graphs $\Gamma(n, m)$ a Kinoshita–Wolcott graph?

(4) Are there infinitely many braids A such that the graph $G(A, 0, 0)$ is a Brunnian theta graph which is not a Kinoshita–Wolcott graph? We conjecture the answer to be yes.

(5) For what $A \in \ker \phi$ and $t_1, t_2 \in \mathbb{Z}$ is $G(A, t_1, t_2)$ a *hyperbolic* Brunnian theta graph? We conjecture that whenever $G(A, t_1, t_2)$ is a Brunnian theta graph, then it is hyperbolic.

(6) Is it true that if $G(A, t_1, t_2)$ is hyperbolic then $G(A^n, t_1, t_2)$ is hyperbolic for all $n \in \mathbb{N}$? Does the hyperbolic volume of the exterior of $G(A^n, t_1, t_2)$ grow linearly in n?

Acknowledgements

We thank the attendees at the 2013 Spatial Graphs conference for helpful discussions, particularly Erica Flapan and Danielle O'Donnol. We are also grateful to Ryan Blair, Ilya Kofman, Jessica Purcell, and Maggy Tomova for helpful conversations. This research was partially funded by Colby College.

References

[Calcut and Metcalf-Burton 0] J. S. Calcut and J. R. Metcalf-Burton, "Double branched covers of theta-curves", *J. Knot Theory Ramifications* **0**:0 (0), 1650046.

[Conway 1970] J. H. Conway, "An enumeration of knots and links, and some of their algebraic properties", pp. 329–358 in *Computational problems in abstract algebra* (Oxford, 1967), edited by J. Leech, Pergamon, Oxford, 1970. MR 0258014 Zbl 0202.54703

[Cromwell 2004] P. R. Cromwell, *Knots and links*, Cambridge University Press, 2004. MR 2107964 Zbl 1066.57007

[Eudave-Muñoz 1992] M. Eudave-Muñoz, "Band sums of links which yield composite links. The cabling conjecture for strongly invertible knots", *Trans. Amer. Math. Soc.* **330**:2 (1992), 463–501. MR 1112545 Zbl 0778.57004

[Gordon 2009] C. Gordon, "Dehn surgery and 3-manifolds", pp. 21–71 in *Low dimensional topology*, edited by T. S. Mrowka and P. S. Ozsváth, IAS/Park City Mathematics Series **15**, American Mathematical Society, Providence, RI, 2009. MR 2503492 Zbl 1194.57003

[Heard 2013] D. Heard, "Orb", 2013, http://www.ms.unimelb.edu.au/~snap/orb.html.

[Kauffman and Lambropoulou 2003] L. H. Kauffman and S. Lambropoulou, "On the classification of rational knots", *Enseign. Math.* (2) **49**:3-4 (2003), 357–410. MR 2028021 Zbl 1059.57004

[Kinoshita 1958] S. Kinoshita, "Alexander polynomials as isotopy invariants. I", *Osaka Math. J.* **10** (1958), 263–271. MR 0102819 Zbl 0119.38801

[Kinoshita 1972] S. Kinoshita, "On elementary ideals of polyhedra in the 3-sphere", *Pacific J. Math.* **42** (1972), 89–98. MR 0312485 Zbl 0239.55002

[Litherland 1989] R. Litherland, "The Alexander module of a knotted theta-curve", *Math. Proc. Cambridge Philos. Soc.* **106**:1 (1989), 95–106. MR 994083 Zbl 0705.57005

[Livingston 1995] C. Livingston, "Knotted symmetric graphs", *Proc. Amer. Math. Soc.* **123**:3 (1995), 963–967. MR 1273507 Zbl 0829.57004

[McAtee et al. 2001] J. McAtee, D. S. Silver, and S. G. Williams, "Coloring spatial graphs", *J. Knot Theory Ramifications* **10**:1 (2001), 109–120. MR 1822144 Zbl 0998.57009

[Montesinos 1975] J. M. Montesinos, "Surgery on links and double branched covers of S^3", pp. 227–259 in *Knots, groups, and* 3*-manifolds: Papers dedicated to the memory of R. H. Fox*, edited by L. P. Neuwirth, Annals of Mathematics Studies **84**, Princeton University Press, 1975. MR 0380802 Zbl 0325.55004

[Motohashi 2000] T. Motohashi, "2-bridge θ-curves in S^3", *Topology Appl.* **108**:3 (2000), 267–276. MR 1794559

[Scharlemann 1992] M. Scharlemann, "Some pictorial remarks on Suzuki's Brunnian graph", pp. 351–354 in *Topology '90* (Columbus, OH, 1990), edited by B. Apanasov et al., Ohio State University Mathematical Research Institute Publications **1**, de Gruyter, Berlin, 1992. MR 1184420 Zbl 0772.57005

[Scharlemann and Thompson 1994] M. Scharlemann and A. Thompson, "Thin position and Heegaard splittings of the 3-sphere", *J. Differential Geom.* **39**:2 (1994), 343–357. MR 1267894 Zbl 0820.57005

[Schubert 1956] H. Schubert, "Knoten mit zwei Brücken", *Math. Z.* **65** (1956), 133–170. MR 0082104 Zbl 0071.39002

[Simon and Wolcott 1990] J. K. Simon and K. Wolcott, "Minimally knotted graphs in S^3", *Topology Appl.* **37**:2 (1990), 163–180. MR 1080350 Zbl 0723.57008

[Thurston 1997] W. P. Thurston, *Three-dimensional geometry and topology*, vol. 1, edited by S. Levy, Princeton Mathematical Series **35**, Princeton University Press, 1997. MR 1435975 Zbl 0873.57001

[Thurston 2002] D. P. Thurston, "The algebra of knotted trivalent graphs and Turaev's shadow world", pp. 337–362 in *Invariants of knots and 3-manifolds* (Kyoto, 2001), edited by T. Ohtsuki et al., Geometry & Topology Monographs **4**, (GTP), Coventry, 2002. MR 2048108 Zbl 1163.57301

[Wolcott 1987] K. Wolcott, "The knotting of theta curves and other graphs in S^3", pp. 325–346 in *Geometry and topology: manifolds, varieties, and knots* (Athens, GA, 1985), edited by C. McCrory and T. Shifrin, Lecture Notes in Pure and Applied Mathematics **105**, Dekker, New York, NY, 1987. MR 873302 Zbl 0613.57003

Received: 2015-08-17 Revised: 2015-10-20 Accepted: 2015-10-22

byoungwookjang@gmail.com *Department of Mathematics and Statistics, Colby College c/o Scott Taylor, Waterville, ME 04901-8858, United States*

abkronau@colby.edu *Department of Mathematics and Statistics, Colby College, Waterville, ME 04901-8858, United States*

luitelpratap@gmail.com *Department of Mathematics and Statistics, Colby College c/o Scott Taylor, Waterville, ME 04901-8858, United States*

demedici@colby.edu *Department of Mathematics and Statistics, Colby College, Waterville, ME 04901-8858, United States*

scott.taylor@colby.edu *Department of Mathematics and Statistics, Colby College, 5832 Mayflower Hill, Waterville, ME 04901-8858, United States*

zupan@unl.edu *Department of Mathematics, University of Nebraska-Lincoln, 1400 R Street, Lincoln, NE 68588, United States*

mathematical sciences publishers msp

INVOLVE 9:5 (2016)
dx.doi.org/10.2140/involve.2016.9.877

Some nonsimple modules for centralizer algebras of the symmetric group

Craig Dodge, Harald Ellers, Yukihide Nakada and Kelly Pohland

(Communicated by Kenneth S. Berenhaut)

James classified the simple modules over the group algebra $k\Sigma_n$ using modules denoted D^λ, where λ is a partition of n. In particular, he showed that D^λ is simple or zero for every partition λ and, furthermore, that for every simple $k\Sigma_n$-module S there exists a partition λ such that $D^\lambda \cong S$. This paper is an extension of a paper of Dodge and Ellers in which they studied analogous modules $\mathcal{D}^{(\lambda,\mu)}$ over the centralizer algebra $k\Sigma_n^{\Sigma_l}$, where λ is a partition of n and μ a partition of l. For every positive prime p we find counterexamples to their conjecture that the $k\Sigma_n^{\Sigma_l}$-modules $\mathcal{D}^{(\lambda,\mu)}$ are always simple or zero, where k is a field of characteristic p. We also study the relationship between $\mathcal{D}^{(\lambda,\mu)}$ and $\mathrm{Hom}_{k\Sigma_l}(D^\mu, \mathrm{res}^{\Sigma_n}_{\Sigma_l} D^\lambda)$ in special cases.

1. Introduction

Let n be a positive integer and k an algebraically closed field of characteristic p. James [1978] studied simple modules over the group algebra $k\Sigma_n$, where Σ_n is the symmetric group on n letters. He defined for each partition $\lambda \vdash n$ the permutation module M^λ with basis consisting of all λ-tabloids. The *Specht module* S^λ is defined to be the submodule of M^λ generated by polytabloids. The kernel intersection theorem can be used to characterize S^λ as

$$S^\lambda = \bigcap \{\ker\varphi \mid \varphi : M^\lambda \to M^{\lambda'},\ \lambda' \rhd \lambda\},$$

where $\lhd$ is the dominance order on partitions [James 1998, p. 97]. He also defined a bilinear form on M^λ using the set of tabloids as an orthonormal basis and proved in [James 1998, 2.2] using the characterization of S^λ above that

$$S^{\lambda\perp} = \sum \{\mathrm{im}\,\varphi \mid \varphi : M^{\lambda'} \to M^\lambda,\ \lambda' \rhd \lambda\}.$$

MSC2010: 20C05, 20C20.
Keywords: centralizer algebras, symmetric groups, modular representations.

James then defined the module D^λ by

$$D^\lambda = S^\lambda/(S^\lambda \cap S^{\lambda\perp})$$

and proved that D^λ is always zero or simple, that $D^\lambda \neq 0$ if and only if λ is p-regular, and that all simple $k\Sigma_n$-modules occur exactly once as λ runs through all p-regular partitions.

Dodge and Ellers applied similar ideas to study representations of centralizer algebras of the symmetric group. In general, let G be a finite group, let H be a subgroup of G, and let k be an algebraically closed field of characteristic p. The centralizer algebra kG^H is defined by

$$kG^H = \{a \in kG \mid ah = ha,\ \forall h \in H\}.$$

Given a kG-module M and a kH-module N we can construct a kG^H-module in a very natural way. The space

$$\operatorname{Hom}_{kH}(N, \operatorname{res}^G_H M)$$

can be given a natural action by kG^H in the following manner:

$$(a\varphi)(t) = a(\varphi(t))$$

for all $a \in kG^H$, $t \in N$ and $\varphi \in \operatorname{Hom}_{kH}(N, \operatorname{res}^G_H M)$.

Dodge and Ellers [2016] studied the representation theory of $k\Sigma_n{}^{\Sigma_l}$, where Σ_n is the symmetric group on n letters, $l \leq n$, and Σ_l is identified with a subgroup of Σ_n permuting the first l letters. Here we review the notation and definitions they used. Let $\mu \vdash l$ and $\lambda \vdash n$. Define a dominance relation on such partition pairs (λ, μ) by

$$(\lambda', \mu') \rhd (\lambda, \mu) \quad \text{if } \lambda' \rhd \lambda \text{ or } (\lambda' = \lambda \text{ and } \mu' \rhd \mu).$$

Define the $k\Sigma_n^{\Sigma_l}$-module

$$\mathcal{M}^{(\lambda,\mu)} = (M^\mu, M^\lambda).$$

This module is designed to be analogous to the permutation modules of the symmetric group. They then define the modules

$$\begin{aligned}
\mathcal{S}^{(\lambda,\mu)} &= \bigcap \{\ker\varphi \mid \varphi : \mathcal{M}^{(\lambda,\mu)} \to \mathcal{M}^{(\lambda',\mu')},\ (\lambda', \mu') \rhd (\lambda, \mu)\},\\
\mathcal{S}^{(\lambda,\mu)\perp} &= \sum \{\operatorname{im}\varphi \mid \varphi : \mathcal{M}^{(\lambda',\mu')} \to \mathcal{M}^{(\lambda,\mu)},\ (\lambda', \mu') \rhd (\lambda, \mu)\},\\
\mathcal{D}^{(\lambda,\mu)} &= \mathcal{S}^{(\lambda,\mu)}/(\mathcal{S}^{(\lambda,\mu)} \cap \mathcal{S}^{(\lambda,\mu)\perp}).
\end{aligned}$$

In the above definitions φ is a $k\Sigma_n^{\Sigma_l}$-module homomorphism. Note that a bilinear form on $\mathcal{M}^{\lambda,\mu}$ has not been defined; the notation for the module $\mathcal{S}^{(\lambda,\mu)\perp}$ was chosen to highlight its similarity to $S^{\lambda\perp}$ in [James 1978]. Paralleling the approach to the representation theory of $k\Sigma_n$ in [James 1978], Dodge and Ellers [2016] proved that if $\lambda \vdash n$ and $\mu \vdash l$, and $l < p$, then $\mathcal{D}^{(\lambda,\mu)}$ is either simple or zero, in agreement with James' result. In addition, they showed that

$$\mathcal{D}^{(\lambda,\mu)} \cong \mathrm{Hom}_{k\Sigma_l}(D^{\mu}, \mathrm{res}_{\Sigma_l}^{\Sigma_n} D^{\lambda})$$

under the same conditions. They conjectured that these facts hold in general when λ and μ are p-regular. In this paper we compute explicit examples to test their conjectures.

For all positive prime p, we explicitly compute the structures of

$$\mathrm{Hom}_{\Sigma_p}(D^{(p)}, \mathrm{res}_{\Sigma_p}^{\Sigma_{p+3}} D^{(p+2,1)})$$

in Sections 3, 4, and 5 and the structures of $\mathcal{D}^{((p+2,1),(p))}$ in Sections 6 and 7. In particular, we show that the space $\mathrm{Hom}_{\Sigma_p}(D^{(p)}, \mathrm{res}_{\Sigma_p}^{\Sigma_{p+3}} D^{(p+2,1)})$ is neither simple nor zero and prove the following characterizations of $\mathcal{D}^{((p+2,1),(p))}$:

Proposition 1.1. *Let k be a field of characteristic p, where $p \neq 3$. Then*

$$\mathcal{D}^{((p+2,1),(p))} \cong \mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}_{\Sigma_p}^{\Sigma_{p+3}} D^{(p+2,1)})$$

as $k\Sigma_{p+3}^{\Sigma_p}$-modules and therefore $\mathcal{D}^{((p+2,1),(p))}$ is neither simple nor zero.

Proposition 1.2. *Let k be a field of characteristic 3. Then*

$$\mathcal{D}^{((5,1),(3))} \cong \mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)})/L$$

as $k\Sigma_6^{\Sigma_3}$-modules, where L is a submodule isomorphic to $\mathcal{M}^{((6),(3))}$. Moreover, $\mathcal{D}^{((5,1),(3))}$ is neither simple nor zero.

Thus neither is simple nor zero for any characteristic p, contrary to the conjectures of Dodge and Ellers. In addition, this shows that the isomorphism conjectured above does not hold in characteristic 3. Finally, in Section 9 we show that in characteristic 2 there is no ordering on pairs of partitions for which the conjectures hold when $n = 5$ and $l = 2$.

2. $\mathcal{M}^{((p+3),\mu)}$ in arbitrary characteristic p

We consider the relationship between the spaces $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}_{\Sigma_p}^{\Sigma_{p+3}} D^{(p+2,1)})$ and $\mathcal{D}^{((p+2,1),(p))}$ when p is a positive prime. Since the pairs of partitions (λ, μ) such that $(\lambda, \mu) \rhd ((p+2, 1), (p))$ are those of the form $((p+3), \mu)$, where $\mu \vdash p$, we first study the modules corresponding to such pairs.

Proposition 2.1. *Let k be a field of characteristic p. Then all modules of the form $\mathcal{M}^{((p+3),\mu)}$, where $\mu \vdash p$, are one-dimensional and mutually isomorphic as $k\Sigma_{p+3}^{\Sigma_p}$-modules.*

Proof. Fix a partition $\mu \vdash p$ and a nonzero tabloid $y_0 \in M^\mu$. From [James 1978, Theorem 13.19] we know that $\mathcal{M}^{((p+3),\mu)}$ is nonzero, so we may choose a nonzero $f \in \mathcal{M}^{((p+3),\mu)}$. Since $f(y_0) \in M^{(p+3)} \cong k$, we have

$$f(y_0) = \sigma f(y_0) = f(\sigma y_0)$$

for all $\sigma \in \Sigma_p$, and since M^μ is a cyclic $k\Sigma_p$-module generated by any nonzero tabloid, it follows that $f(y) = f(y_0)$ for any tabloid $y \in M^\mu$. Thus if $f_0 \in \mathcal{M}^{((p+3),\mu)}$ is defined by $f_0(y_0) = 1$ then $\mathcal{M}^{((p+3),\mu)} = \mathrm{span}\{f_0\}$ as a $k\Sigma_{p+3}^{\Sigma_p}$-module. In particular, it is one-dimensional.

We now describe a generating set for $k\Sigma_{p+3}^{\Sigma_p}$. From [Kleshchev 2005, Proposition 2.1.1] we have

$$k\Sigma_{p+3}^{\Sigma_p} = \langle Z(k\Sigma_p),\ (p{+}1\ \ p{+}2),\ (p{+}1\ \ p{+}2\ \ p{+}3),\ L_{p+1},\ L_{p+2},\ L_{p+3}\rangle,$$

where $Z(k\Sigma_p)$ is the center of $k\Sigma_p$ and L_k is the *Jucys–Murphy element* defined as

$$L_k = \sum_{1 \le m < k} (m\ k).$$

It is well known that $Z(k\Sigma_p)$ is spanned by elements $s_\tau \in k\Sigma_p$ for τ a partition of p, where s_τ denotes the sum of all elements in Σ_p with cycle type corresponding to the partition τ. Let K_τ denote the conjugacy class corresponding to the partition τ. Since any element of Σ_{p+3} acts trivially on the codomain of $\mathrm{Hom}_{k\Sigma_p}(D^\mu, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+3)}) = \mathcal{M}^{((p+3),\mu)}$, we deduce that the action of the module is described by the table

	f_0
s_τ	$\lvert K_\tau\rvert f_0$
$(p{+}1\ \ p{+}2)$	f_0
$(p{+}1\ \ p{+}2\ \ p{+}3)$	f_0
L_{p+1}	0
L_{p+2}	f_0
L_{p+3}	$2f_0$

Since our choice of μ was arbitrary, it follows that all modules of the form $\mathcal{M}^{((p+3),\mu)}$ are mutually isomorphic, as claimed. □

3. $\mathrm{Hom}_{k\Sigma_2}(D^{(2)}, \mathrm{res}^{\Sigma_5}_{\Sigma_2} D^{(4,1)})$ in characteristic 2

Next we determine the structure of $\mathrm{Hom}_{k\Sigma_2}(D^{(2)}, \mathrm{res}^{\Sigma_5}_{\Sigma_2} D^{(4,1)})$. In this and all following sections, when $D^\lambda \cong S^\lambda$ we will identify a coset in D^λ with its corresponding element in S^λ as an abuse of notation. We first note that $D^{(2)}$ is trivial by definition.

We have that $M^{(4,1)}$ is spanned by

$$\left\{ \frac{\overline{2\ 3\ 4\ 5}}{\underline{1}}, \frac{\overline{1\ 3\ 4\ 5}}{\underline{2}}, \frac{\overline{1\ 2\ 4\ 5}}{\underline{3}}, \frac{\overline{1\ 2\ 3\ 5}}{\underline{4}}, \frac{\overline{1\ 2\ 3\ 4}}{\underline{5}} \right\}.$$

We will denote these tabloids by x_1, x_2, x_3, x_4, x_5, respectively. Since $x_2, \dots, x_5$ correspond to the standard tableau in $M^{(4,1)}$, we know from [James 1978, Theorem 8.4] that the Specht module $S^{(4,1)}$ has basis $\{x_2-x_1,\ x_3-x_1,\ x_4-x_1,\ x_5-x_1\}$. For simplicity we denote each element in this basis by $c_i = x_i - x_1$ for $2 \le i \le 5$. To compute $S^{(4,1)\perp}$, note that since the map $\mathcal{M}^{((5),(2))} \to \mathcal{M}^{((4,1),(2))}$ defined by $1 \mapsto x_1 + x_2 + x_3 + x_4 + x_5$ is a $k\Sigma_5^{\Sigma_2}$-module homomorphism, it follows that $x_1+x_2+x_3+x_4+x_5 \in S^{(4,1)\perp}$. Moreover, since $S^{(4,1)}$ is four-dimensional we can conclude from [James 1978, 1.3] that $S^{(4,1)\perp}$ has dimension 1 and hence that $S^{(4,1)\perp}$ has basis $\{x_1+x_2+x_3+x_4+x_5\}$. Notice that $S^{\lambda} \cap S^{\lambda\perp} = 0$, so $D^{(4,1)} \cong S^{(4,1)}$. Now, fix $z \in D^{(2)}$ with $z \neq 0$, and let

$$f : D^{(2)} \to \operatorname{res}^{\Sigma_5}_{\Sigma_2} D^{(4,1)}$$

be defined by

$$f(z) = a_2c_2 + a_3c_3 + a_4c_4 + a_5c_5.$$

Observe that since $D^{(2)} \cong k$ and k is a field of characteristic 2, we have $f \in \operatorname{Hom}_{k\Sigma_2}(D^{(2)}, \operatorname{res}^{\Sigma_5}_{\Sigma_2} D^{(4,1)})$ if and only if $[(1)+(12)]f = 0$. Therefore, we need

$$\begin{aligned} [(1)+(12)]f(z) &= a_2c_2 + a_3c_3 + a_4c_4 + a_5c_5 - a_2c_2 \\ &\qquad\qquad + a_3(c_3-c_2) + a_4(c_4-c_2) + a_5(c_5-c_2) \\ &= -a_3c_2 - a_4c_2 - a_5c_2 \\ &= -(a_3+a_4+a_5)c_2 = 0. \end{aligned}$$

Thus f is a $k\Sigma_2$-module homomorphism exactly when $a_3 + a_4 + a_5 = 0$. Hence f has the form

$$\begin{aligned} f(z) &= a_2c_2 + a_3c_3 + a_4c_4 + (-a_3-a_4)c_5 \\ &= a_2c_2 + a_3(c_3-c_5) + a_4(c_4-c_5). \end{aligned}$$

Therefore a basis for $\operatorname{Hom}_{k\Sigma_2}(D^{(2)}, \operatorname{res}^{\Sigma_5}_{\Sigma_2} D^{(4,1)})$ is

$$\alpha(z) = c_2 = x_1 + x_2, \quad \beta(z) = c_3 - c_5 = x_3 - x_5, \quad \gamma(z) = c_4 - c_5 = x_4 - x_5.$$

Next we examine how $k\Sigma_5^{\Sigma_2}$ acts on $\{\alpha, \beta, \gamma\}$. As our generators for $k\Sigma_5^{\Sigma_2}$, we will be using the generating set from Proposition 2.1, namely

$$k\Sigma_5^{\Sigma_2} = \langle (1),\ (12),\ (34),\ (345),\ L_3,\ L_4,\ L_5 \rangle.$$

The action of the module is described by the table

	α	β	γ
(12)	α	β	γ
(34)	α	γ	β
(345)	α	$\gamma-\beta$	$-\beta$
L_3	α	α	0
L_4	0	γ	$\alpha+\beta$
L_5	0	$-\alpha-\gamma$	$-\alpha-\beta$

Thus we can see that span$\{\alpha\}$ is a submodule of $\mathrm{Hom}_{k\Sigma_2}(D^{(2)}, \mathrm{res}_{\Sigma_2}^{\Sigma_5} D^{(4,1)})$. Comparing this table with that on page 880 describing $\mathcal{M}^{((5),(2))}$, we see that span$\{\alpha\} \not\cong \mathcal{M}^{((5),(2))}$. The quotient by this one-dimensional submodule has basis $\{\bar{\beta}, \bar{\gamma}\}$, and the action of the module is described by the table

	$\bar{\beta}$	$\bar{\gamma}$
(12)	$\bar{\beta}$	$\bar{\gamma}$
(34)	$\bar{\gamma}$	$\bar{\beta}$
(345)	$\bar{\gamma}-\bar{\beta}$	$-\bar{\beta}$
L_3	$\bar{0}$	$\bar{0}$
L_4	$\bar{\gamma}$	$\bar{\beta}$
L_5	$-\bar{\gamma}$	$-\bar{\beta}$

We will show that this is a simple two-dimensional module. If this is not simple, it must contain a one-dimensional submodule. We leave to the reader the easy confirmation that span$\{\bar{\beta}\}$ and span$\{\bar{\gamma}\}$ are not submodules. So suppose $a_0, a_1 \neq 0$ and assume for contradiction that the one-dimensional k-vector space span$\{a_0\bar{\beta}+a_1\bar{\gamma}\}$ is a submodule. It follows then that

$$((34)+(345))(a_0\bar{\beta}+a_1\bar{\gamma}) \in \mathrm{span}\{a_0\bar{\beta}+a_1\bar{\gamma}\},$$

so we have

$$\begin{aligned}((34)+(345))(a_0\bar{\beta}+a_1\bar{\gamma}) &= a_0(34)\bar{\beta}+a_1(34)\bar{\gamma}+a_0(345)\bar{\beta}+a_1(345)\bar{\gamma}\\ &= a_0\bar{\gamma}+a_1\bar{\beta}+a_0\bar{\gamma}-a_0\bar{\beta}-a_1\bar{\beta}\\ &= -a_0\bar{\beta}.\end{aligned}$$

Thus, it must be that $a_0 = 0$, a contradiction. Thus, for all $a_0, a_1 \in k$, we have that span$\{a_0\bar{\beta}+a_1\bar{\gamma}\}$ is not a submodule of $\mathrm{Hom}_{k\Sigma_2}(D^{(2)}, \mathrm{res}_{\Sigma_2}^{\Sigma_5} D^{(4,1)})/\,\mathrm{span}\{\alpha\}$, so the quotient is a two-dimensional simple module.

4. $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)})$ in characteristic 3

Let k be a field of characteristic 3. We now determine the structure of

$$\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)}).$$

Notice that $D^{(3)}$ is trivial. We have that $M^{(5,1)}$ is spanned by

$$\left\{ \frac{\overline{2\ 3\ 4\ 5\ 6}}{\underline{1}}, \frac{\overline{1\ 3\ 4\ 5\ 6}}{\underline{2}}, \frac{\overline{1\ 2\ 4\ 5\ 6}}{\underline{3}}, \frac{\overline{1\ 2\ 3\ 5\ 6}}{\underline{4}}, \frac{\overline{1\ 2\ 3\ 4\ 6}}{\underline{5}}, \frac{\overline{1\ 2\ 3\ 4\ 5}}{\underline{6}} \right\}.$$

We will again denote these standard tabloids by x_1, x_2, x_3, x_4, x_5, x_6, respectively. Since $x_2, \ldots, x_6$ correspond to the standard tableau in $M^{(5,1)}$, we know from [James 1978, Theorem 8.4] that the Specht module $S^{(5,1)}$ is spanned by $\{x_2-x_1,\ x_3-x_1,\ x_4-x_1,\ x_5-x_1,\ x_6-x_1\}$. For simplicity we denote each element in this basis by $c_i = x_i - x_1$ for $2 \le i \le 6$. To compute $S^{(5,1)\perp}$, note that since the map $\mathcal{M}^{((6),(2))} \to \mathcal{M}^{((5,1),(2))}$ defined by $1 \mapsto x_1+x_2+x_3+x_4+x_5+x_6$ is a $k\Sigma_6^{\Sigma_3}$-module homomorphism, it follows that $x_1+x_2+x_3+x_4+x_5+x_6 \in S^{(5,1)\perp}$. Moreover, since $S^{(5,1)}$ is five-dimensional, we can conclude from [James 1978, 1.3] that $S^{(5,1)\perp}$ has dimension 1 and hence that $S^{(5,1)\perp}$ has basis $\{x_1+x_2+x_3+x_4+x_5+x_6\}$. From this, it is clear that $S^{(5,1)} \cap S^{(5,1)\perp} = 0$, so $D^{(5,1)} \cong S^{(5,1)}$. We now fix $z \in D^{(3)}$ with $z \neq 0$ and let

$$f : D^{(3)} \to \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)}$$

be defined by

$$f(z) = a_2c_2 + a_3c_3 + a_4c_4 + a_5c_5 + a_6c_6.$$

Since Σ_3 is generated by (12) and (13), we have $f \in \mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)})$ exactly when $f(z) = (12)f(z)$ and $f(z) = (13)f(z)$. Thus we must have

$$\begin{aligned}(12)f(z) &= a_2(-c_2) + a_3(c_3-c_2) + a_4(c_4-c_2) + a_5(c_5-c_2) + a_6(c_6-c_2)\\ &= (-a_2-a_3-a_4-a_5-a_6)c_2 + a_3c_3 + a_4c_4 + a_5c_5 + a_6c_6,\end{aligned}$$

so $a_2 = -a_2-a_3-a_4-a_5-a_6$. Similarly,

$$\begin{aligned}(13)f(z) &= a_2(c_2-c_3) + a_3(-c_3) + a_4(c_4-c_3) + a_5(c_5-c_3) + a_6(c_6-c_3)\\ &= a_2c_2 + (-a_2-a_3-a_4-a_5-a_6)c_3 + a_4c_4 + a_5c_5 + a_6c_6,\end{aligned}$$

so $a_3 = -a_2-a_3-a_4-a_5-a_6$. Thus $a_2 = a_3$, and since $a_2 = -a_2-a_3-a_4-a_5-a_6$ and k has characteristic 3, we get that $0 = a_4+a_5+a_6$. Consequently,

$$f(z) = a_2c_2 + a_3c_3 + a_4c_4 + a_5c_5 + a_6c_6 = a_2(c_2+c_3) + a_4(c_4-c_6) + a_5(c_5-c_6).$$

Therefore, we get that $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}_{\Sigma_3}^{\Sigma_6} D^{(5,1)})$ is spanned by $\{\alpha, \beta, \gamma\}$, where

$$\alpha(z) = c_2+c_3 = x_1+x_2+x_3, \quad \beta(z) = c_4-c_6 = x_4-x_6, \quad \gamma(z) = c_5-c_6 = x_5-x_6.$$

The table describing the action on this basis is

	α	β	γ
$(12)+(13)+(23)$	0	0	0
$(123)+(132)$	2α	2β	2γ
(45)	α	γ	β
(456)	α	$2\beta+\gamma$	2β
L_4	2α	α	0
L_5	0	γ	$\alpha+\beta$
L_6	α	$2\alpha+2\gamma$	$2\alpha+2\beta$

From this table we can deduce that $\operatorname{span}\{\alpha\}$ and $\operatorname{span}\{\alpha+\beta+\gamma\}$ are submodules of $\operatorname{Hom}_{k\Sigma_3}(D^{(3)}, \operatorname{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})$. The table describing the action on $\operatorname{span}\{\alpha+\beta+\gamma\}$ is

	$\alpha+\beta+\gamma$
$(12)+(13)+(23)$	0
$(123)+(132)$	$2(\alpha+\beta+\gamma)$
(45)	$\alpha+\beta+\gamma$
(456)	$\alpha+\beta+\gamma$
L_4	0
L_5	$\alpha+\beta+\gamma$
L_6	$2(\alpha+\beta+\gamma)$

Comparing these tables to that on page 880, we see that $\operatorname{span}\{\alpha\} \ncong \mathcal{M}^{((6),(3))}$ and $\operatorname{span}\{\alpha+\beta+\gamma\} \cong \mathcal{M}^{((6),(3))}$. The corresponding quotient

$$\operatorname{Hom}_{k\Sigma_3}(D^{(3)}, \operatorname{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})/(\operatorname{span}\{\alpha\} \oplus \operatorname{span}\{\alpha+\beta+\gamma\})$$

is one-dimensional with basis $\{\bar{\beta}\}$ and the table describing the action on this basis is

	$\bar{\beta}$
$(12)+(13)+(23)$	$\bar{0}$
$(123)+(132)$	$2\bar{\beta}$
(45)	$2\bar{\beta}$
(456)	$\bar{\beta}$
L_4	$\bar{0}$
L_5	$2\bar{\beta}$
L_6	$\bar{\beta}$

Note that $\{\bar{\beta}\}$ is isomorphic to neither $\operatorname{span}\{a\}$ nor $\mathcal{M}^{((6),(3))}$.

5. $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$ in characteristic p

Let $p \geq 5$ be prime, and let k be a field of characteristic p. We determine the structure of

$$\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)}).$$

Notice that $D^{(p)}$ is trivial. Using notation similar to that in Sections 3 and 4, $M^{(p+2,1)}$ is spanned by $\{x_1, \dots, x_{p+3}\}$. From computations entirely analogous to those in characteristics 2 and 3, we know that the Specht module $S^{(p+2,1)}$ has basis $\{c_2, \dots, c_{p+3}\}$, where $c_i = x_i - x_1$ for $2 \leq i \leq p+3$, and that $S^{(p+2,1)\perp}$ has dimension 1 with basis $\{x_1+x_2+\cdots+x_{p+3}\}$. Consequently $S^{(p+2,1)} \cap S^{(p+2,1)\perp} = 0$ and $D^{(p+2,1)} \cong S^{(p+2,1)}$.

Fix $z \in D^{(p)}$ with $z \neq 0$. Let $f : D^{(p)} \to \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)}$ be defined by $f(z) = a_2c_2 + a_3c_3 + \cdots + a_{p+3}c_{p+3}$. Notice that

$$f \in \mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$$

if and only if $f(z) = (12)f(z) = (13)f(z) = \cdots = (1\ p)f(z)$ since Σ_p is generated by $(12), \dots, (1\ p)$. Since

$$(1\ i)f(z) = a_2(c_2 - c_i) + a_3(c_3 - c_i) + \cdots + a_i(-c_i) + \cdots + a_{p+3}(c_{p+3} - c_i)$$

and

$$f(z) = a_2c_2 + a_3c_3 + \cdots + a_{p+3}c_{p+3},$$

it follows that for all $2 \leq i \leq p$ we must have

$$\begin{aligned} a_2(c_2 - c_i) + a_3(c_3 - c_i) + \cdots + a_i(-c_i) + \cdots + a_{p+3}(c_{p+3} - c_i)& \\ = a_2c_2 + a_3c_3 + \cdots + a_{p+3}c_{p+3}.& \end{aligned}$$

Simplifying, we have

$$a_i c_i = \left(-\sum_{k=2}^{p+3} a_k\right) c_i,$$

so $a_i = -a_2 - a_3 - \cdots - a_{p+3}$. Since this holds for arbitrary $2 \leq i \leq p$, we get that $a_2 = a_3 = \cdots = a_p$. In particular, substituting this into the above equality with $i = 2$ we have

$$a_{p+1} + a_{p+2} + a_{p+3} = -a_2 - a_2 - a_3 - \cdots - a_p = -pa_2 = 0$$

since k has characteristic p. Hence f must have the form

$$\begin{aligned} f(z) &= a_2c_2 + a_3c_3 + \cdots + a_{p+3}c_{p+3} \\ &= a_2(c_2 + c_3 + \cdots + c_p) + a_{p+1}(c_{p+1} - c_{p+3}) + a_{p+2}(c_{p+2} - c_{p+3}). \end{aligned}$$

From this, we can see that a basis for $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$ is $\{\alpha, \beta, \gamma\}$, where

$$\alpha(z) = c_2 + \cdots + c_p = x_1 + x_2 + \cdots + x_p,$$
$$\beta(z) = c_{p+1} - c_{p+3} = x_{p+1} - x_{p+3},$$
$$\gamma(z) = c_{p+2} - c_{p+3} = x_{p+2} - x_{p+3}.$$

Recall from Proposition 2.1 that for a partition τ, we let s_τ denote the sum of all elements in Σ_p with cycle type corresponding to τ and let K_τ denote the conjugacy class corresponding to τ. Notice that since each element of Σ_p permutes $\{1, \dots, p\}$, we can conclude that $\sigma\alpha = \alpha$, $\sigma\beta = \beta$, and $\sigma\gamma = \gamma$ for any $\sigma \in \Sigma_p$. From this we can derive the action of $k\Sigma_{p+1}^{\Sigma_p}$ on this basis, and the table describing this is

	α	β	γ
s_τ	$\lvert K_\tau\rvert\alpha$	$\lvert K_\tau\rvert\beta$	$\lvert K_\tau\rvert\gamma$
$(p{+}1\;\; p{+}2)$	α	γ	β
$(p{+}1\;\; p{+}2\;\; p{+}3)$	α	$\gamma-\beta$	$-\beta$
L_{p+1}	$-\alpha$	α	0
L_{p+2}	0	γ	$\alpha+\beta$
L_{p+3}	α	$-\alpha-\gamma$	$-\alpha-\beta$

Notice that span$\{\alpha\}$ is a submodule. Comparing its action to the action described in the table on page 880 we see that span$\{\alpha\} \ncong \mathcal{M}^{((p+3),(p))}$. The table describing the action on the corresponding quotient module is

	$\bar\beta$	$\bar\gamma$
s_τ	$\lvert K_\tau\rvert\bar\beta$	$\lvert K_\tau\rvert\bar\gamma$
$(p{+}1\;\; p{+}2)$	$\bar\gamma$	$\bar\beta$
$(p{+}1\;\; p{+}2\;\; p{+}3)$	$\bar\gamma-\bar\beta$	$-\bar\beta$
L_{p+1}	$\bar0$	$\bar0$
L_{p+2}	$\bar\gamma$	$\bar\beta$
L_{p+3}	$-\bar\gamma$	$-\bar\beta$

We now show that this quotient is simple. Since the quotient is two-dimensional, we can show that it is simple by showing that there are no one-dimensional submodules. We leave it to the reader to confirm that span$\{\bar\beta\}$ and span$\{\bar\gamma\}$ are not submodules. So let $a_0, a_1 \neq 0$ and suppose for contradiction that span$\{a_0\bar\beta + a_1\bar\gamma\}$ is a submodule. Then

$$\begin{aligned}\big((p{+}1\;\; p{+}2)+(p{+}1\;\; p{+}2\;\; p{+}3)\big)(a_0\bar\beta+a_1\bar\gamma) &= a_0\bar\gamma+a_1\bar\beta+a_0(\bar\gamma-\bar\beta)+a_1(-\bar\beta)\\ &= 2a_0\bar\gamma-a_0\bar\beta,\end{aligned}$$

so for some $c \in k$, we have $ca_0 = -a_0$ and $ca_1 = 2a_0$. Thus, $c = -1$ and $a_1 = -2a_0$. Similarly,

$$\big((p+1\ \ p+2)-(p+1\ \ p+2\ \ p+3)\big)(a_0\bar{\beta}+a_1\bar{\gamma}) = a_0\bar{\gamma}+a_1\bar{\beta}-a_0(\bar{\gamma}-\bar{\beta})-a_1(-\bar{\beta})$$
$$= (a_0+2a_1)\bar{\beta},$$

so $a_0+2a_1 = 0$ since $a_1 \neq 0$. Thus, $a_0 = -2a_1$, and since $a_1 = -2a_0$, we must have $a_1 = 4a_1$. For char $k \neq 3$ this is a contradiction. Hence, $\operatorname{span}\{a_0\bar{\beta}+a_1\bar{\gamma}\}$ is not a submodule for all $a_0, a_1 \in k$ and the quotient is a two-dimensional simple module.

6. $\mathcal{D}^{((p+2,1),(p))}$ in characteristic $p \neq 3$

In this section we compute the structure of $\mathcal{D}^{((p+2,1),(p))}$ over a field of characteristic p when $p \neq 3$ and prove Proposition 1.1. To compute the structure of $\mathcal{D}^{((p+2,1),(p))}$ we will need the following lemma.

Lemma 6.1. *Let A be a finite-dimensional k-algebra, let $S_1, \dots, S_n$ be simple A-modules, and suppose K and L are A-modules with L having no S_i as a composition factor and K having every S_i as a composition factor. Let $\varphi : K \to L$ be an A-module homomorphism, and let M be minimal among submodules of K having every S_i as a composition factor. Then $M \subseteq \ker\varphi$.*

Proof. Suppose, for contradiction, that $M \not\subseteq \ker\varphi$. Then the inclusion $M \supset \ker\varphi \cap M$ is strict. Refine the filtration $M \supset (\ker\varphi \cap M) \supseteq 0$ into a composition series. Since M is minimal among submodules of K having every S_i as a composition factor, they cannot all belong to the composition series of $\ker\varphi \cap M$. Thus S_1, without loss of generality, is a composition factor of $M/(\ker\varphi \cap M)$. But

$$M/(\ker\varphi \cap M) \cong \varphi(M) \subseteq L,$$

so S_1 is a composition factor of L, a contradiction. □

The remainder of this section will be devoted to the proof of Proposition 1.1.

Suppose k has characteristic $p \neq 3$. We first compute a basis for $\mathcal{M}^{((p+2,1),(p))}$. For each $1 \leq i \leq p+3$, let t_i be the $(p+2, 1)$-tableau with i in the second row, and let $x_i = \{t_i\}$. Then $\{x_1, \dots, x_{p+3}\}$ forms a basis for $M^{(p+2,1)}$.

Let $0 \neq z \in M^{(p)}$ and let $f : M^{(p)} \to \operatorname{res}^{\Sigma_{p+3}}_{\Sigma_p} M^{(p+2,1)}$ be defined by

$$f(z) = \sum_{n=1}^{p+3} a_n x_n.$$

Since the transpositions $(1\ \ i)$ for $1 \leq i \leq p$ generate the group Σ_p, for f to be a Σ_p-homomorphism it is sufficient that $[(1)-(1\ \ i)]f = 0$ for all $2 \leq i \leq p$. Fix one such i. Then

$$\begin{aligned}[(1)-(1\ \ i)]f(z) &= \left(\sum_{n=1}^{p+3} a_n x_n\right) - \left(a_i x_1 + a_1 x_i + \sum_{n\neq 1,i} a_n x_n\right)\\ &= a_1x_1 + a_i x_i - a_i x_1 - a_1 x_i\\ &= (a_1 - a_i)(x_1 - x_i).\end{aligned}$$

Thus we must have $a_1 = a_i$. Since this must be true for all $2 \le i \le p$, we deduce that $\mathcal{M}^{((p+2,1),(p))}$ has a basis $\{\alpha, \beta'_{p+1}, \beta'_{p+2}, \beta'_{p+3}\}$, where

$$\begin{aligned} \alpha(z) &= x_1 + \cdots + x_p, & \beta'_{p+2}(z) &= x_{p+2},\\ \beta'_{p+1}(z) &= x_{p+1}, & \beta'_{p+3}(z) &= x_{p+3}.\end{aligned}$$

From this it is easy to check that

$$\begin{aligned} \alpha(z) &= x_1 + \cdots + x_p, & \beta_{p+2}(z) &= x_{p+2} - x_{p+3},\\ \beta_{p+1}(z) &= x_{p+1} - x_{p+3}, & \beta_{p+3}(z) &= x_1 + \cdots + x_{p+3}\end{aligned}$$

is also a basis for $\mathcal{M}^{((p+2,1),(p))}$. The set $\{\alpha, \beta_{p+1}, \beta_{p+2}\}$ can be identified with the basis of $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$ found in Section 5, so we can deduce that

$$N = \mathrm{span}\{\alpha, \beta_{p+1}, \beta_{p+2}\}$$

is a subspace of $\mathcal{M}^{((p+2,1),(p))}$ isomorphic to $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, \mathrm{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$. Furthermore, the table describing the action on β_{p+3} is

	β_{p+3}
s_τ	$\lvert K_\tau\rvert \beta_{p+3}$
$(p{+}1\ \ p{+}2)$	β_{p+3}
$(p{+}1\ \ p{+}2\ \ p{+}3)$	β_{p+3}
L_{p+1}	0
L_{p+2}	β_{p+3}
L_{p+3}	$2\beta_{p+3}$

so $K = \mathrm{span}\{\beta_{p+3}\}$ is a submodule of $\mathcal{M}^{((p+2,1),(p))}$, and comparing this table to that on page 880 we see that it is isomorphic to $\mathcal{M}^{((p+3),(p))}$. Hence we have the direct sum decomposition

$$\mathcal{M}^{((p+2,1),(p))} = N \oplus K.$$

We now compute $\mathcal{D}^{((p+2,1),(p))}$. Since we know from Section 5 that the composition factors of $\mathrm{Hom}_{k\Sigma_p}(D^{(p)}, D^{(p+2,1)})$ consist of simple modules not isomorphic to $\mathcal{M}^{((p+3),(p))}$, it follows from Lemma 6.1 that $N \subseteq \ker\varphi$ for every

$\varphi : \mathcal{M}^{((p+2,1),(p))} \to \mathcal{M}^{((p+3),(p))}$, so that $N \subseteq \mathcal{S}^{((p+2,1),(p))}$. The reverse inclusion follows from the fact that N is the kernel of the projection of $\mathcal{M}^{((p+2,1),(p))}$ onto $K \cong \mathcal{M}^{((p+3),(p))}$. Hence

$$\mathcal{S}^{((p+2,1),(p))} = N.$$

We can deduce that $K \subseteq \mathcal{S}^{((p+2,1),(p))\perp}$ since K is the image of the map

$$\mathcal{M}^{((p+3),(p))} \to \mathcal{M}^{((p+2,1),(p))}$$

consisting of the isomorphism to K followed by injection. For the reverse inclusion, let $\varphi : \mathcal{M}^{((p+3),(p))} \to \mathcal{M}^{((p+2,1),(p))}$ be nonzero. Since $\operatorname{im}\varphi \cong \mathcal{M}^{((p+3),(p))}$ by Schur's lemma and K is the only composition factor of $\mathcal{M}^{((p+2,1),(p))}$ isomorphic to $\mathcal{M}^{((p+3),(p))}$, we must have $\operatorname{im}\varphi = K$. Consequently $K \subseteq \mathcal{S}^{((p+2,1),(p))\perp}$ by definition. Thus

$$\mathcal{S}^{((p+2,1),(p))\perp} = K.$$

Since $K \cap N = \{0\}$, we have

$$\mathcal{D}^{((p+2,1),(p))} = \mathcal{S}^{((p+2,1),(p))}/\{0\} \cong N \cong \operatorname{Hom}_{k\Sigma_p}(D^{(p)}, \operatorname{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$$

as claimed. We showed in Sections 3 and 5 that $\operatorname{Hom}_{k\Sigma_p}(D^{(p)}, \operatorname{res}^{\Sigma_{p+3}}_{\Sigma_p} D^{(p+2,1)})$ was neither simple nor zero for $p \neq 3$, and so the same must be true of $\mathcal{D}^{((p+2,1),(p))}$.

7. $\mathcal{D}^{((5,1),(3))}$ in characteristic 3

In this section we compute the structure of $\mathcal{D}^{((5,1),(3))}$ over a field of characteristic 3 and prove Proposition 1.2. This module has a structure different from the analogous modules $\mathcal{D}^{((p+2,1),(p))}$ in other characteristics because the spanning set $\{\alpha, \beta_{p+1}, \beta_{p+2}, \beta_{p+3}\}$ in $\mathcal{M}^{((p+2,1),(p))}$ fails to be linearly independent in characteristic 3. The remainder of this section will be devoted to the proof of Proposition 1.2.

The method used in the proof of Proposition 1.1 to find a basis for $\mathcal{M}^{((p+2,1),(p))}$ works when $p = 3$, so we have a basis

$$\alpha(z) = x_1 + x_2 + x_3, \quad \beta'_4(z) = x_4, \quad \beta'_5(z) = x_5, \quad \beta'_6(z) = x_6$$

for $\mathcal{M}^{((5,1),(3))}$. However, since

$$\begin{aligned}(x_1 + x_2 + x_3) + (x_4 - x_6) + (x_5 - x_6) &= x_1 + x_2 + x_3 + x_4 + x_5 - 2x_6 \\ &= x_1 + x_2 + x_3 + x_4 + x_5 + x_6\end{aligned}$$

in characteristic 3, the set $\{\alpha, \beta_{p+1}, \beta_{p+2}, \beta_{p+3}\}$ used for the characteristic $p \neq 3$ case in Section 6 fails to be independent. Thus we use the basis

$$\alpha(z) = x_1 + x_2 + x_3, \quad \beta_4(z) = x_4 - x_6, \quad \beta_5(z) = x_5 - x_6, \quad \gamma_6(z) = x_6.$$

The set $\{\alpha, \beta_4, \beta_5\}$ can be identified with the basis of $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})$ found in Section 4. Thus we can deduce that $N = \mathrm{span}\{\alpha, \beta_4, \beta_5\}$ is a submodule of $\mathcal{M}^{((5,1),(3))}$ isomorphic to $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})$. The corresponding quotient has basis $\{\overline{\gamma_6}\}$ and the table describing the action on this basis is

	$\overline{\gamma_6}$
$(12)+(13)+(23)$	0
$(123)+(132)$	$2\overline{\gamma_6}$
(45)	$\overline{\gamma_6}$
(456)	$\overline{\gamma_6}$
L_4	0
L_5	$\overline{\gamma_6}$
L_6	$2\overline{\gamma_6}$

Comparing this table with that on page 880 we see that

$$\mathrm{span}\{\overline{\beta_6}\} \cong \mathcal{M}^{((6),(3))}.$$

We now compute $\mathcal{D}^{((5,1),(3))}$. Recall that $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})$ has two composition factors S_1 and S_2 not isomorphic to $\mathcal{M}^{((6),(3))}$, so that the same is true of N. Since N is the kernel of the projection

$$\mathcal{M}^{((5,1),(3))} \to \mathcal{M}^{((5,1),(3))}/N \cong \mathcal{M}^{((6),(3))},$$

we have $\mathcal{S}^{((5,1),(3))} \subseteq N$. To show the reverse inclusion, fix a homomorphism

$$\varphi : \mathcal{M}^{((5,1),(3))} \to \mathcal{M}^{((6),\mu)},$$

where $\mu \vdash 3$; by Proposition 2.1 we know that $\mathcal{M}^{((6),\mu)} \cong \mathrm{res}^{k\Sigma_6}_{k\Sigma_6^{\Sigma_3}} k$. Suppose $\varphi(\beta_4') = a$. Then

$$\begin{aligned}
\varphi(\alpha) &= \varphi(L_4 x) = L_4 a = 0,\\
\varphi(\beta_4) &= \varphi((1-(46))\beta_4') = (1-(46))a = 0,\\
\varphi(\beta_5) &= \varphi(((45)-(46))\beta_4') = ((45)-(46))a = 0,
\end{aligned}$$

so $\varphi(N) = 0$. Thus $N \subseteq \ker\varphi$, and since our choice of φ was arbitrary, it follows that $N \subseteq \mathcal{S}^{((5,1),(3))}$. Consequently

$$\mathcal{S}^{((5,1),(3))} = N.$$

From Section 4 we know that $\mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})$ has a submodule L isomorphic to $\mathcal{M}^{((6),(3))}$. Since

$$N \cong \mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)}),$$

it follows that N also has a corresponding submodule K isomorphic to $\mathcal{M}^{((6),(3))}$. We can deduce that $K \subseteq \mathcal{S}^{((5,1),(3))\perp}$ since K is the image of the map

$$\mathcal{M}^{((6),(3))} \to \mathcal{M}^{((5,1),(3))}$$

consisting of the isomorphism to K followed by injection. Since the image of any homomorphism $\mathcal{M}^{((6),(3))} \to \mathcal{M}^{((5,1),(3))}$ must be isomorphic to $\mathcal{M}^{((6),(3))}$ and the only composition factor of N isomorphic to $\mathcal{M}^{((6),(3))}$ is K, it follows that

$$\mathcal{S}^{((5,1),(3))} \cap \mathcal{S}^{((5,1),(3))\perp} = K.$$

Thus

$$\mathcal{D}^{((5,1),(3))} = N/K \cong \mathrm{Hom}_{k\Sigma_3}(D^{(3)}, \mathrm{res}^{\Sigma_6}_{\Sigma_3} D^{(5,1)})/L$$

as claimed.

8. $\mathcal{M}^{(\lambda,\mu)}$ for $\lambda \vdash 5,\ \mu \vdash 2$

The above computations show that in every positive characteristic there are pairs of partitions (λ, μ) for which $\mathcal{D}^{(\lambda,\mu)}$ is neither simple nor zero, as conjectured in [Dodge and Ellers 2016]. However, it may be the case that this may be fixed by choosing a different ordering on pairs of partitions; that is, it may be the case that there exists a different ordering on pairs of partitions for which $\mathcal{D}^{(\lambda,\mu)}$ is always simple or zero. In this section we use the computer algebra system Magma [Bosma et al. 1997] to generate the structure of the $k\Sigma_5^{\Sigma_2}$-module $M^{(\lambda,\mu)}$ when $\lambda \vdash 5$ and $\mu \vdash 2$, and in the next section use this information to show that there does not exist any such ordering in characteristic 2.

We will treat the cases when $\mu = (2)$ and $\mu = (1^2)$ separately.

Case 1: $\mu = (1^2)$. Since $M^{(1^2)} \cong k\Sigma_2$ as $k\Sigma_2$-modules, we have

$$\mathcal{M}^{(\lambda,1^2)} = \mathrm{Hom}_{k\Sigma_2}(k\Sigma_2, \mathrm{res}^{\Sigma_5}_{\Sigma_2} M^{\lambda}) \cong \mathrm{res}^{\Sigma_5}_{\Sigma_2} M^{\lambda}$$

so we may compute in M^{λ}. This can be defined in Magma as a $k\Sigma_5$-module through the command

K := PermutationModule(Sym(5), YoungSubgroup(λ : Full := 5), GF(2));

However, we wish to define $\mathcal{M}^{(\lambda,1^2)}$ as a $k\Sigma_5^{\Sigma_2}$-module. To do this we will find the matrices of the action of the generators of $k\Sigma_5^{\Sigma_2}$ on the basis of M^{λ}, and then create a module over the matrix algebra that they generate.

Given an $x \in k\Sigma_5^{\Sigma_2}$ we may find the matrix of x acting on the basis of M^{λ} through the function

```
mapmatrix := func<x | Matrix(GF(2), Dimension(K), Dimension(K),
  [(VectorSpace(GF(2), Dimension(K)) ! (K.i * x)) : i in {1..Dimension(K)}])>;
```

This function simply creates the matrix of x in the natural way. Magma has a default basis for K, namely the elements $K.i$ for $1 \leq i \leq \dim K$. Thus, for the i-th basis vector $K.i$ of K, we find $K.i * x$ in terms of the basis of K and set it as the i-th row of the matrix.

We will be using the generating set for $k\Sigma_5^{\Sigma_2}$ given in Section 3, namely

$$k\Sigma_5^{\Sigma_2} = \langle (12),\ (34),\ (345),\ L_3,\ L_4,\ L_5 \rangle.$$

Using the function *mapmatrix* we can create the matrix algebra generated by the matrices of the actions of these generators through the command

```
A := MatrixAlgebra<GF(2), Dimension(K) |
  mapmatrix((Sym(5) ! (1,2))),
  mapmatrix((Sym(5) ! (3,4))),
  mapmatrix((Sym(5) ! (3,4,5))),
  mapmatrix((Sym(5) ! (1,3))) + mapmatrix((Sym(5) ! (2,3))),
  mapmatrix((Sym(5) ! (1,4))) +
    mapmatrix((Sym(5) ! (2,4))) + mapmatrix((Sym(5) ! (3,4))),
  mapmatrix((Sym(5) ! (1,5))) + mapmatrix((Sym(5) ! (2,5))) +
    mapmatrix((Sym(5) ! (3,5))) + mapmatrix((Sym(5) ! (4,5))) >;
```

We can then generate $\mathcal{M}^{(\lambda,1^2)}$ as a $k\Sigma_5^{\Sigma_2}$-module through the command

```
M := RModule(A);
```

and find its constituents with multiplicities via

```
ConstituentsWithMultiplicities(M);
```

Case 2: $\mu = (2)$. We first find a basis for $\mathcal{M}^{(\lambda,(2))}$.

Proposition 8.1. *Suppose k is a field of characteristic 2, let $\lambda \vdash 5$, and fix a nonzero $z \in M^{(2)} \cong k$. Then the functions defined by*

$$f_x(z) = \begin{cases} x + (12)x & \text{if } x \neq (12)x, \\ x & \text{if } x = (12)x, \end{cases}$$

where x is a λ-tabloid, constitute a basis for $\mathcal{M}^{(\lambda,(2))}$.

Proof. The independence of the functions f_x follows immediately from the independence of the tableau in M^{λ}. Fix a nonzero $z \in M^{(2)} \cong k$ and let

$$f : M^{(2)} \to \operatorname{res}_{\Sigma_2}^{\Sigma_5} M^{\lambda}$$

be defined by

$$f(z) = \sum_{x \text{ a } \lambda\text{-tabloid}} a_x x.$$

To have $f \in \mathcal{M}^{(\lambda,(2))}$ it is necessary and sufficient that $[(1)-(12)]f(z)=0$. Thus we need

$$0=[(1)-(12)]f(z)=\sum_{x \text{ a } \lambda\text{-tabloid}} a_x x-\sum_{x \text{ a } \lambda\text{-tabloid}} a_x(12)x$$
$$=\sum_{x \text{ a } \lambda\text{-tabloid}} a_x x-\sum_{x \text{ a } \lambda\text{-tabloid}} a_{(12)x}x=\sum_{x \text{ a } \lambda\text{-tabloid}} (a_x-a_{(12)x})x.$$

Thus we must have $a_x=a_{(12)x}$ for all x. This means that $f(z)$ is a linear combination of the functions $f_x(z)$, as needed. □

As before, we generate $K=M^\lambda$ as a permutation module over $k\Sigma_5$. To find a basis for $\mathcal{M}^{(\lambda,(2))}$ we first create a list consisting of sums of elements which are mapped to each other via the transposition (12). We accomplish this through the procedure below:

```
BasisSet := [];
BasisGen := procedure(~BasisSet, K)
  for i in {1..Dimension(K)} do
      if K.i + K.i * (Sym(5) ! (1,2)) eq Zero(K) then
              Append(~BasisSet, K.i);
      elif K.i + K.i * (Sym(5) ! (1,2)) in BasisSet then
              print "Skip";
      else
              Append(~BasisSet, K.i + K.i * (Sym(5) ! (1,2)));
      end if;
    end for;
end procedure;
BasisGen(~BasisSet, K);
```

For every basis element $K.i$ of K, we add $K.i((1)+(1,2))$ to the list BasisGen of basis elements if $K.i((1)+(1,2))$ is nonzero and $K.i$ if it is zero. This constitutes a basis for $\mathcal{M}^{(\lambda,(2))}$ by Proposition 8.1. The elif statement excludes duplicate basis elements.

Having created a list of basis elements for $\mathcal{M}^{(\lambda,(2))}$, we create the space spanned by them as a subspace of the vector space of appropriate dimension. We can do this through

```
W := sub<VectorSpace(GF(2), Dimension(K)) | [Eltseq(s) : s in BasisSet]>;
```

The Eltseq command coerces each basis element into a tuple so that it can be embedded into the vector space.

Although our basis vectors are now elements of a vector space and not a permutation module, we can still act on them by elements of $k\Sigma_5$ by coercing vectors in W back into M^λ. We exploit this property to find the matrix of the action of generators of $k\Sigma_5^{\Sigma_2}$ on $\mathcal{M}^{(\lambda,(2))}$ as follows:

```
A := MatrixAlgebra<GF(2), Dimension(W) |
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (1,2))))
    : i in {1..Dimension(W)}]),
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (3,4))))
    : i in {1..Dimension(W)}]),
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (3,4,5))))
    : i in {1..Dimension(W)}]),
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (1,3)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (2,3))))
    : i in {1..Dimension(W)}]),
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (1,4)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (2,4)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (3,4))))
    : i in {1..Dimension(W)}]) >;
Matrix(GF(2), Dimension(W), Dimension(W), [Coordinates(W, W ! (
    (K ! Basis(W)[i]) * (Sym(5) ! (1,5)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (2,5)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (3,5)) +
    (K ! Basis(W)[i]) * (Sym(5) ! (4,5))))
    : i in {1..Dimension(W)}]) >;
```

The principle is identical to the algorithm used in the case $\mu = (1^2)$. The only difference is that we are working in the intermediary vector space W rather than directly in M^λ.

Having generated the algebra, we can define the desired module and find its constituents with multiplicities as before.

9. Alternative partial orders

The structures of $\mathcal{M}^{(\lambda,\mu)}$ when $\lambda \vdash 5$ and $\mu \vdash 2$ are compiled in the Appendix. The key piece of information we will use is that $\mathcal{M}^{(\lambda,\mu)}$ has at least three composition factors, except when $(\lambda, \mu) = ((5), (2))$ or $(\lambda, \mu) = ((5), (1^2))$, in which case we have $\mathcal{M}^{((5),(2))} \cong \mathcal{M}^{((5),(1^2))}$ and both are one-dimensional. In particular, when

(λ, μ) is neither $((5), (2))$ nor $((5), (1^2))$, we know that $\mathcal{M}^{(\lambda,\mu)}$ has two composition factors nonisomorphic to $\mathcal{M}^{((5),(2))}$.

Using this fact, we prove the following:

Proposition 9.1. *In characteristic* 2, *there exists no ordering on pairs of partitions* (λ, μ) *for which* $\mathcal{D}^{(\lambda,\mu)}$ *is always simple or zero.*

Proof. Let $\rhd$ be an arbitrary total order on pairs of partitions and let (λ_0, μ_0) be the most dominant partition such that (λ_0, μ_0) is not $((5), (2))$ or $((5), (1^2))$. If (λ_0, μ_0) is the most dominant partition then by definition $\mathcal{D}^{(\lambda_0,\mu_0)} \cong \mathcal{S}^{(\lambda_0,\mu_0)} = \mathcal{M}^{(\lambda_0,\mu_0)}$, so $\mathcal{D}^{(\lambda,\mu)}$ is neither simple nor zero. Otherwise (λ_0, μ_0) is dominated by $((5), (2))$ or $((5), (1^2))$ or both. Then since $\mathcal{M}^{(\lambda_0,\mu_0)}$ has two composition factors not isomorphic to $\mathcal{M}^{((5),(2))} \cong \mathcal{M}^{((5),(1^2))}$, it follows from [Dodge and Ellers 2016, 1.2] that $\mathcal{D}^{(\lambda_0,\mu_0)}$ has two composition factors not isomorphic to $\mathcal{M}^{((5),(2))} \cong \mathcal{M}^{((5),(1^2))}$. In particular it is neither simple nor zero, as claimed. □

10. Concluding remarks

In Sections 6 and 7 we showed that the conjecture that $\mathcal{D}^{(\lambda,\mu)}$ is always simple or zero fails in every positive characteristic p, while Section 9 shows that in general a different choice of partial orders will not correct the conjecture. However, in every example computed in this paper $\mathcal{D}^{(\lambda,\mu)}$ has had at most two composition factors, and they have always been distinct. This suggests that there may still be a bound on the composition length of $\mathcal{D}^{(\lambda,\mu)}$, even if it is not one as conjectured by Dodge and Ellers.

In [Danz et al. 2013], Danz, Ellers, and Murray answered in the negative the question of whether the FG^H-module $\mathrm{Hom}_{FH}(S, \mathrm{res}^G_H T)$ is always simple or zero for G a finite group and H a subgroup, F a field of positive characteristic, S a simple FH-module, and T a simple FG-module. However, it was still open whether there were counterexamples when FG and FH were symmetric group algebras. Our computations in Sections 3, 4, and 5 provided examples of spaces of the form $\mathrm{Hom}_{k\Sigma_l}(D^\mu, \mathrm{res}^{\Sigma_n}_{\Sigma_l} D^\lambda)$ which were neither simple nor zero, answering this question in the negative as well. The space described in Section 4 has also provided a counterexample to the conjecture that $\mathcal{D}^{(\lambda,\mu)} \cong \mathrm{Hom}_{k\Sigma_l}(D^\mu, \mathrm{res}^{\Sigma_n}_{\Sigma_l} D^\lambda)$ when $\mu \vdash l$ and $\lambda \vdash n$, as demonstrated in Section 7. However, unlike the conjecture on the simplicity of $\mathcal{D}^{(\lambda,\mu)}$, we have only been able to provide a counterexample in characteristic 3: the computations in Section 6 are in agreement with the conjecture. Although we have shown that isomorphism cannot hold in general, it may be the case that $\mathcal{D}^{(\lambda,\mu)}$ is always isomorphic to a quotient of $\mathrm{Hom}_{k\Sigma_l}(D^\mu, \mathrm{res}^{\Sigma_n}_{\Sigma_l} D^\lambda)$.

Finally, Dodge and Ellers [2016] established that every simple $k\Sigma_n^{\Sigma_l}$-module appears as a composition factor of some $\mathcal{D}^{(\lambda,\mu)}$. Though we have shown that those simple modules are not the modules $\mathcal{D}^{(\lambda,\mu)}$ themselves, our calculations may give hints as to how the simple modules appear as composition factors of the $\mathcal{D}^{(\lambda,\mu)}$. In particular, in our calculations the modules $\mathcal{D}^{(\lambda,\mu)}$ always have a simple head. Thus it is possible that the simple modules appear as simple heads of the $\mathcal{D}^{(\lambda,\mu)}$, in the same way that the simple $k\Sigma_n$-modules D^λ appear as the simple heads of the Specht module S^λ when λ is p-regular.

Appendix: $\mathcal{M}^{(\lambda,\mu)}$ when $\lambda \vdash 5$ and $\mu \vdash 2$

$\mathcal{M}^{(\lambda,\mu)}$	d	Multiplicity
$\mathcal{M}^{((5),(2))}$	1	1
$\mathcal{M}^{((5),(1^2))}$	1	1
$\mathcal{M}^{((4,1),(2))}$	1	1
	1	1
	2	2
$\mathcal{M}^{((4,1),(1^2))}$	1	2
	1	1
	2	1
$\mathcal{M}^{((3,1,1),(2))}$	1	2
	1	3
	2	2
	2	2
$\mathcal{M}^{((3,1,1),(1^2))}$	1	4
	1	4
	2	2
	2	4
$\mathcal{M}^{((2,2,1),(2))}$	1	2
	1	4
	2	4
	2	2

$\mathcal{M}^{(\lambda,\mu)}$	d	Multiplicity
$\mathcal{M}^{((2,2,1),(1^2))}$	1	4
	1	6
	2	8
	2	2
$\mathcal{M}^{((2,1,1,1),(2))}$	1	4
	1	7
	2	3
	2	8
$\mathcal{M}^{((2,1,1,1),(1^2))}$	1	8
	1	12
	2	4
	2	16
$\mathcal{M}^{((1^5),(2))}$	1	12
	1	8
	2	16
	2	4
$\mathcal{M}^{((1^5),(1^2))}$	1	16
	1	24
	2	8
	2	32

Table 1. The constituents of $\mathcal{M}^{(\lambda,\mu)}$ are modules of dimension d (given in the middle column) over GF(2) with corresponding multiplicities given in the third column.

Acknowledgments

Yukihide Nakada was supported in full by Allegheny College's Harold M. State Research Fellowship. Kelly Pohland was supported in full by Allegheny College's Dr. Barbara Lotze Student-Faculty Research Fellowship Fund. The authors gratefully acknowledge the helpful input of an anonymous referee.

References

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR 1484478 Zbl 0898.68039

[Danz et al. 2013] S. Danz, H. Ellers, and J. Murray, "The centralizer of a subgroup in a group algebra", *Proc. Edinb. Math. Soc.* (2) **56**:1 (2013), 49–56. MR 3021404 Zbl 1272.20002

[Dodge and Ellers 2016] C. J. Dodge and H. Ellers, "Specht modules and simple modules for centralizer algebras of the symmetric group", *Comm. Algebra* **44**:2 (2016), 837–850. MR 3449956 Zbl 06573329

[James 1978] G. D. James, *The representation theory of the symmetric groups*, Lecture Notes in Mathematics **682**, Springer, Berlin, 1978. MR 513828 Zbl 0393.20009

[James 1998] G. James, "Symmetric groups and Schur algebras", pp. 91–102 in *Algebraic groups and their representations* (Cambridge, 1997), edited by R. W. Carter and J. Saxl, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **517**, Kluwer Acad. Publ., Dordrecht, 1998. MR 1670766 Zbl 0917.20009

[Kleshchev 2005] A. Kleshchev, *Linear and projective representations of symmetric groups*, Cambridge Tracts in Mathematics **163**, Cambridge University Press, 2005. MR 2165457 Zbl 1080.20011

Received: 2015-10-16 Revised: 2016-02-15 Accepted: 2016-05-13

cdodge2@allegheny.edu *Department of Mathematics, Allegheny College, 520 North Main St., Meadville, PA 16335, United States*

hellers@allegheny.edu *Department of Mathematics, Allegheny College, 520 North Main St., Meadville, PA 16335, United States*

yukihide.nakada@gmail.com *Department of Mathematics, Allegheny College, 520 North Main St., Meadville, PA 16335, United States*

pohlandk@allegheny.edu *Department of Mathematics, Allegheny College, 520 North Main St., Meadville, PA 16335, United States*

PACIFIC JOURNAL OF MATHEMATICS
Vol. 9, No. 5, 2016
dx.doi.org/10.2140/pjm.2016.9.899

ACKNOWLEDGEMENT

The editors gratefully acknowledge the valuable advice of the referees who helped them select and better the papers appearing in *Involve, a Journal of Mathematics* (reports dated January 1, 2015 through June 30, 2016).

Colin Adams, Sukumar Das Adhikari, Ayman Badawi, Christian Ballot, John Baxley, Thomas Bell, Catherine Anne Bénéteau, Paul Bezandry, Albrecht Boettcher, Joseph E. Borzellino, Mark Broom, Scott T. Chapman, Hara Charalambous, Angel Chavez, Goong Chen, David Clampitt, James Cossey, Annalisa Crannell, Jonathan Cutler, Samantha Dahlberg, Oliver Dasbach, Jonathan Dawes, Nedim Degirmenci, Joe DeMaio, William Devanny, William Dickinson, Michael Dorff, Steven Dougherty, Gregory Dresden, Hendrikus Duifhuis, Behrouz Emamizadeh, John Engbers, Dana C. Ernst, Michael Filaseta, Thomas Fleming, Joseph A. Gallian, Martino Garonzi, Frank G. Garvan, Severino Gervacio, Ronald Gould, Philip Jameson Graber, Richard M. Green, Jerrold Griggs, Helen Grundman, Maria Pia Gualdani, Jim Haglund, Heekyoung Hahn, Christopher Hammond, Marshall Hampton, Jane M. Hawkins, Allison Henrich, Julian Hook, Natalia Hritonenko, John Hutchens, Gizem Karaali, Dagan Karp, Safeer Hussain Khan, William Kinnersley, Richard Klima, William Klostermeyer, Lori Koban, Lingju Kong, Sergey Korotov, Jakub Kozik, Michal Krizek, Dominic Andrew Lanphier, Gary Lawlor, Arne Ledet, Mark L. Lewis, Cong Li, Xiyue Liao, Der-Fen Liu, Xueyan Sherry Liu, Lewis D. Ludwig, Vince Lyzinski, Colton Magnant, John Maharry, Ana Martínez-Pastor, Blake Mellor, Mary C. Meyer, Brittney Miller, Robert Moss, John Murray, Kellen Myers, Kieka Mynhardt, Sivaram K. Narayan, Shahla Nasserasr, Sam Nelson, Carlo Nitsch, James O'Shea, Joseph petrillo, Christopher Phan, Paolo Piccione, Pietro Poggi-Corradini, Mario Ponce, Rachel Pries, Anna Pun, Christophe Reutenauer, Brendon Rhoades, Tim R. Riley, Gareth E. Roberts, Jeremy Rouse, Heather M. Russell, Shin Satoh, Felix Schlenk, John Schmitt, Sam Scholze, Daniel Schultheis, Andrew Schultz, Mohamed Seoud, Anastasia Shabanskaya, Patrick D. Shanahan, Hirokazu Shimauchi, Ken W. Smith, Lawrence Somer, Joe Stickles, Paul Stockmeyer, Huy Tran, Lisa Traynor, Serge Eugene Troubetzkoy, Katharine Turner, Torsten Ueckerdt, Vasily Vasyunin, Ashlee Versypt, Xinmao Wang, Elizabeth Wilcox, Shelby Wilson, Matthew Woolf, Carl R. Yerger, Petr Zemanek, Michael E. Zieve.

Guidelines for Authors

Submissions in all mathematical areas are encouraged. All manuscripts accepted for publication in *Involve* are considered publishable in quality journals in their respective fields, and include a minimum of one-third student authorship. Submissions should include substantial faculty input; faculty co-authorship is strongly encouraged. 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, 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@msp.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

2016 vol. 9 no. 5