involve

a journal of mathematics

msp

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.

MANAGING EDITOR

Kenneth S. Berenhaut   Wake Forest University, USA

PRODUCTION

Silvio Levy, Scientific Editor

Cover: Alex Scorpan

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

■ msp

Intrinsically triple-linked graphs in $\mathbb{R}P^3$

Jared Federman, Joel Foisy, Kristin McNamara and Emily Stark

(Communicated by Kenneth S. Berenhaut)

Flapan, Naimi and Pommersheim (2001) showed that every spatial embedding of K_{10}, the complete graph on ten vertices, contains a nonsplit three-component link; that is, K_{10} is *intrinsically triple-linked* in $\mathbb{R}^3$. The work of Bowlin and Foisy (2004) and Flapan, Foisy, Naimi, and Pommersheim (2001) extended the list of known intrinsically triple-linked graphs in $\mathbb{R}^3$ to include several other families of graphs. In this paper, we will show that while some of these graphs can be embedded 3-linklessly in $\mathbb{R}P^3$, the graph K_{10} is intrinsically triple-linked in $\mathbb{R}P^3$.

1. Introduction

There is a classic theory of knots and links in Euclidean 3-space (or the 3-sphere), and, as Manturov [2004] pointed out, there is a sympathetic theory of knots and links in $\mathbb{R}P^3$. Drobotukhina [1990] developed an analog of the Jones polynomial for the case of oriented links in $\mathbb{R}P^3$, and Mroczkowski [2003] described a method to unknot knots and links in $\mathbb{R}P^3$ through an analog of classical knot and link diagrams for knots in $\mathbb{R}^3$. Flapan, Howards, Lawrence, and Mellor [Flapan et al. 2006] investigated intrinsic linking and knotting in arbitrary 3-manifolds. Here, following Bustamente et al. [2009], we use a weaker notion of unlink than was used in [Flapan et al. 2006], and we examine the intrinsic linking properties of graphs embedded in $\mathbb{R}P^3$. In particular, we will examine graphs that contain a three-component nonsplit link in every embedding into $\mathbb{R}P^3$.

Real projective 3-space $\mathbb{R}P^3$ can be obtained from the 3-ball D^3 by identifying opposite points of its boundary; hence, a link in $\mathbb{R}P^3$ consists of a union of arcs and loops so that the endpoints of any arc lie on antipodal boundary points of the 3-ball. We may use ambient isotopy to move all arcs so that their endpoints lie on a fixed great circle, the "equator" of the ball. Therefore, a link may be represented

in $\mathbb{R}P^2$ by its projection onto a 2-disk, D^2, whose boundary is the equator, with antipodal points identified.

Projective space has a nontrivial first homology group, $H_1(\mathbb{R}P^3) \cong \mathbb{Z}_2$. Let g, the cycle consisting of a line in D^3 running between the north and south poles, be the generator of this group. Using crossing changes and ambient isotopy on an $\mathbb{R}P^2$ projection of a knot, Mroczkowski [2003] showed that every knot in $\mathbb{R}P^3$ can be transformed into either the trivial cycle or g. Thus, there are two nonequivalent unknots in $\mathbb{R}P^3$. Cycles that can be unknotted into a cycle homologous to g will be referred to as *1-homologous cycles*. Cycles that can be unknotted into a trivial cycle will be referred to as *0-homologous cycles*.

Following [Bustamante et al. 2009], we say a link in $\mathbb{R}P^3$ is *splittable* if one component can be contained within a 3-ball embedded in $\mathbb{R}P^3$, while the other component lies in the complement of the 3-ball. Otherwise, a link in $\mathbb{R}P^3$ is said to be *nonsplit*. A nonsplit link may be formed in one of three ways in $\mathbb{R}P^3$: by two 0-homologous cycles, by a 0-homologous cycle and a 1-homologous cycle, and by two 1-homologous cycles. Moreover, since a 1-homologous cycle cannot be contained within a ball embedded in $\mathbb{R}P^3$, two disjoint 1-homologous cycles will always form a nonsplit link. In this paper, we will refer to nonsplit linked cycles as *linked cycles* and to an embedded graph as *linked* if it contains a nonsplit link.

A graph H is a *minor* of G if H can be obtained from G through a series of vertex removals, edge removals, or edge contractions. A graph G is said to be *minor-minimal* with respect to property P if G has property P, but no minor of G has property P. The complete set of minor-minimal intrinsically linked graphs in $\mathbb{R}^3$ is given by the Petersen family graphs, including K_6 and the graphs obtained from K_6 by $\Delta-Y$ and $Y-\Delta$ exchanges [Conway and Gordon 1983; Robertson et al. 1995; Sachs 1984]. However, all Petersen family graphs except $K_{4,4}-\{e\}$, where e is an edge, embed linklessly in $\mathbb{R}P^3$, as shown in [Bustamante et al. 2009], a paper which also exhibits 597 graphs that are minor-minimal intrinsically linked in $\mathbb{R}P^3$. The complete set of minor-minimal intrinsically linked graphs in $\mathbb{R}P^3$ is finite [Robertson and Seymour 2004], and remains to be found.

A *nonsplit triple link* is a nonsplit link of three components, which, in an abuse of language, will be referred to as a *triple link* in this paper. An embedding of a graph is *triple-linked* if it contains a nonsplit link of three components, and a graph is *intrinsically triple-linked in X*, a topological space, if every embedding of the graph into X contains a nonsplit triple link.

Conway and Gordon [1983] and Sachs [1983; 1984] proved that K_6 is intrinsically linked in $\mathbb{R}^3$. In contrast, K_6 can be linklessly embedded in $\mathbb{R}P^3$ (see Figure 3). Bustamante et al. [2009] showed that 7 is the smallest n for which K_n is intrinsically linked in $\mathbb{R}P^3$. Flapan, Naimi, and Pommersheim [Flapan et al. 2001a] proved 10 is the smallest n for which K_n is intrinsically triple-linked in $\mathbb{R}^3$. In Section 3, we show

that 10 is also the smallest n for which K_n is intrinsically triple-linked in $\mathbb{R}P^3$. It remains to show whether K_{10} is minor-minimal with respect to triple-linking in $\mathbb{R}P^3$.

In Section 4, we show that two intrinsically triple-linked graphs in $\mathbb{R}^3$ can be embedded 3-linklessly in $\mathbb{R}P^3$ and exhibit two other minor-minimal intrinsically triple-linked graphs in $\mathbb{R}P^3$. A complete set of minor-minimal intrinsically triple-linked graphs in both $\mathbb{R}^3$ and $\mathbb{R}P^3$ remains to be found. Such sets are finite due to the result in [Robertson and Seymour 2004].

2. Definitions and preliminary lemmas

We begin with some elementary definitions and notation. A *graph*, $G = (V, E)$, is a set of *vertices*, $V(G)$, and *edges*, $E(G)$, where an edge is an unordered pair (v_1, v_2) with $v_1, v_2 \in V(G)$. If G is a graph with $v_1, \ldots, v_n \in V(G)$ and $(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n), (v_n, v_1) \in E(G)$, with $v_i \neq v_j$ for all $i \neq j$, then the sequences of vertices $v_1, \ldots, v_n$ and edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n), (v_n, v_1)$ is an *n-cycle* of G, denoted $(v_1, \ldots, v_n)$. In this paper, we also refer to the image of a cycle under an embedding as an *n*-cycle.

If G is a graph and $v_1, \ldots, v_n \in V(G)$, define the *induced subgraph*, $G[v_1, \ldots, v_n]$, to be the subgraph of G with

$$V(G[v_1, \ldots, v_n]) = \{v_1, \ldots, v_n\},$$
$$E(G[v_1, \ldots, v_n]) = \{(v_i, v_j) \in E(G) \mid v_i, v_j \in \{v_1, \ldots, v_n\}\}.$$

The classical notion of linking number extends to links embedded in $\mathbb{R}P^3$. Suppose L and K are two loops embedded in $\mathbb{R}P^3$; orient L and K. At each crossing, assign $+1$ or -1, as drawn in Figure 1. Then the *mod* 2 *linking number of L and K*, $\mathrm{lk}(L, K)$, is the sum of the numbers, $+1$ or -1, at each crossing in the embedding of L and K divided by 2, taken modulo 2. In $\mathbb{R}P^3$, there are five generalized Reidemeister moves [Manturov 2004], which are drawn in Figure 2. As in $\mathbb{R}^3$, one can use Reidemeister moves to justify that mod 2 linking number is well-defined in $\mathbb{R}P^3$. In particular, the mod 2 linking number of a splittable two-component link is 0. However, in $\mathbb{R}P^3$, the mod 2 linking number need not be an integer; for example, two disjoint 1-homologous cycles can have mod 2 linking number $\pm\frac{1}{2}$.

The following lemmas provide us information about carefully chosen induced subgraphs of the graphs we study.

$+1 \qquad -1$
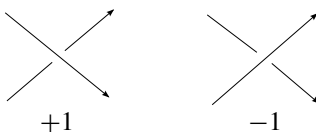
Figure 1. Link crossings.
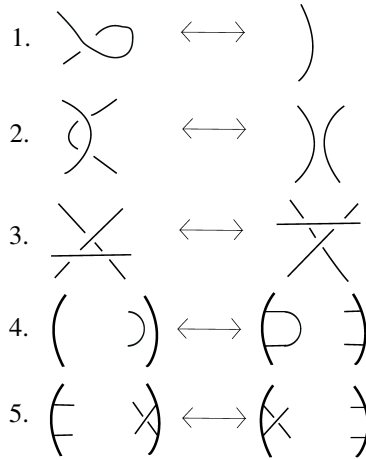
Figure 2. Generalized Reidemeister moves in $\mathbb{R}P^3$.

Lemma 1 [Bustamante et al. 2009]. *The graphs obtained by removing two edges from K_7 and removing one edge from $K_{4,4}$ are intrinsically linked in $\mathbb{R}P^3$.*

Lemma 2 [Bustamante et al. 2009]. *Given a linkless embedding of K_6 in $\mathbb{R}P^3$, no K_4 subgraph can have all 0-homologous cycles.*

In addition, we use the following elementary observation.

Lemma 3. *For every embedding into $\mathbb{R}P^3$, the graph K_4 has an even number of 1-homologous 3-cycles.*

The next two lemmas were shown true in $\mathbb{R}^3$ by [Flapan et al. 2001a] and [Bowlin and Foisy 2004], respectively. In each case, the proof holds analogously in $\mathbb{R}P^3$.

Lemma 4. *Let G be a graph embedded in $\mathbb{R}P^3$ that contains cycles C_1, C_2, C_3 and C_4. Suppose C_1 and C_4 are disjoint from each other and from C_2 and C_3 and suppose $C_2 \cap C_3$ is a simple path. If $\mathrm{lk}(C_1, C_2) \neq 0$ and $\mathrm{lk}(C_3, C_4) \neq 0$, then G contains a nonsplit three-component link.*

Lemma 5. *In an embedded graph with mutually disjoint simple closed curves, $C_1, C_2, C_3,$ and C_4, and two disjoint paths x_1 and x_2 such that x_1 and x_2 begin in C_2 and end in C_3, if $\mathrm{lk}(C_1, C_2) \neq 0$ and $\mathrm{lk}(C_3, C_4) \neq 0$, then the embedded graph contains a nonsplit three-component link.*

3. Intrinsically triple-linked complete graphs on n vertices

The proposition below, that K_{11} is intrinsically triple-linked in $\mathbb{R}P^3$, is not the main result of this paper. In fact, our main result, that K_{10} is intrinsically triple-linked in $\mathbb{R}P^3$, implies this proposition, by a result of [Nešetřil and Thomas 1985]. However,

the proof is included because it is (relatively) concise and follows from examining four carefully chosen subgraphs of K_{11} and applying Lemmas 4 and 5.

Proposition 6. *The graph K_{11} is intrinsically triple-linked in $\mathbb{R}P^3$.*

Proof. Let G be a complete graph isomorphic to K_{11} with vertex set $\{1, 2, \ldots, 11\}$. Embed G in $\mathbb{R}P^3$.

Since K_7 is intrinsically linked in $\mathbb{R}P^3$, the graph $G[1, 2, 3, 4, 5, 6, 7] \cong K_7$ contains a pair of linked cycles. Without loss of generality, suppose the linked cycles are $C_1 = (1, 2, 3)$ and $C_2' = (4, 5, 6, 7)$. Homologically, the cycle $(4, 5, 6, 7)$ is the sum of the cycles $(4, 5, 6)$ and $(4, 6, 7)$. Thus,

$$\text{lk}\big((1, 2, 3), (4, 5, 6, 7)\big) = \text{lk}\big((1, 2, 3), (4, 5, 6)\big) + \text{lk}\big((1, 2, 3), (4, 6, 7)\big).$$

Since the numbers on the right-hand side cannot both equal 0, without loss of generality, $C_1 = (1, 2, 3)$ links with $C_2 = (4, 5, 6)$.

Again, since K_7 is intrinsically linked in $\mathbb{R}P^3$, it follows that the subgraph $G[5, 6, 7, 8, 9, 10, 11] \cong K_7$ contains a pair of linked cycles. In the manner described above, this pair of cycles may be reduced to two linked 3-cycles. If it is not the case that one cycle contains $\{5\}$ and one cycle contains $\{6\}$, then Lemma 4 applies, and G is triple-linked. To handle the other case, suppose, without loss of generality, that $C_3 = (5, 7, 9)$ and $C_4 = (6, 8, 10)$ are the pair of linked cycles in $G[5, 6, 7, 8, 9, 10, 11]$.

To obtain two collections of disjoint 1-homologous cycles, consider two subgraphs isomorphic to K_6. First, if $G[1, 2, 3, 4, 6, 11] \cong K_6$ contains a pair of linked cycles, then one cycle shares vertex $\{6\}$ with C_4 and both are disjoint from C_3, so Lemma 4 applies and G is triple-linked. Otherwise, by Lemma 2, the set

$$A = \{(1, 2, 3), (1, 2, 11), (1, 3, 11), (2, 3, 11)\}$$

contains a 1-homologous cycle, C_5.

Similarly, if $G[6, 7, 8, 9, 10, 11] \cong K_6$ contains a pair of linked cycles, then one cycle shares vertex $\{6\}$ with C_2 and both are disjoint from C_1. So, Lemma 4 applies and G is triple-linked. Otherwise, by Lemma 2, the set

$$B = \{(7, 8, 9), (7, 8, 10), (7, 9, 10), (8, 9, 10)\}$$

contains a 1-homologous cycle, C_6.

Since $A \cap B = \varnothing$, the cycles $C_5 \in A$ and $C_6 \in B$ are disjoint and 1-homologous and hence are linked. So, C_2 and C_6 are disjoint from each other and from C_1 and C_5. In the case that $C_1 = C_5$, the cycles C_1, C_2, and C_6 form a triple link. Otherwise, $C_1 \cap C_5$ is a simple path, so G contains a triple link by Lemma 4. $\square$

To prove that K_{10} is intrinsically triple-linked in $\mathbb{R}P^3$, we first describe how its subgraphs isomorphic to K_6 must be embedded.
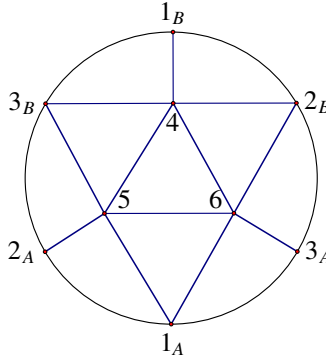
Figure 3. A projection of a linkless embedding of K_6 in $\mathbb{R}P^3$.

Proposition 7. *If G is isomorphic to K_6 and embedded in $\mathbb{R}P^3$ and G contains two disjoint 0-homologous cycles, then G contains a nonsplit link.*

Proof. Let G be isomorphic to K_6 and suppose G is embedded so that it contains two disjoint 0-homologous cycles and no nonsplit link. Without loss of generality, let $(1, 2, 3)$ and $(4, 5, 6)$ be 0-homologous cycles in G. Consider $G[1, 2, 3, 4]$. Since G is not linked, by Lemmas 2 and 3, the subgraph $G[1, 2, 3, 4]$ contains two 1-homologous cycles. Without loss of generality, let $(1, 2, 4)$ and $(1, 3, 4)$ be 1-homologous cycles.

Similarly, $G[2, 4, 5, 6]$ contains two 1-homologous cycles. The cycle $(4, 5, 6)$ is 0-homologous by assumption and since $(2, 5, 6)$ is disjoint from $(1, 3, 4)$, which is 1-homologous, $(2, 5, 6)$ is 0-homologous since G is assumed to have no nonsplit link. Thus, $(2, 4, 5)$ and $(2, 4, 6)$ are 1-homologous cycles.

In addition, $G[1, 2, 3, 6]$ contains two 1-homologous cycles. Since $(1, 2, 3)$ is 0-homologous by assumption and $(1, 3, 6)$ is disjoint from $(2, 4, 5)$, which is 1-homologous, $(1, 2, 6)$ and $(2, 3, 6)$ are 1-homologous.

Finally, $G[1, 3, 5, 6]$ contains two 1-homologous cycles. But, $(2, 4, 6)$, $(2, 4, 5)$, and $(1, 2, 4)$ are 1-homologous and disjoint from $(1, 3, 5)$, $(1, 3, 6)$, and $(3, 5, 6)$ respectively, a contradiction, since $G[1, 3, 5, 6]$ must contain two 1-homologous cycles. □

Proposition 8. *Up to ambient isotopy and crossing changes, Figure 3 describes the only way to linklessly embed K_6 in $\mathbb{R}P^3$.*

Proof. Let G be a complete graph on vertex set $\{1, 2, 3, 4, 5, 6\}$. Embed G in $\mathbb{R}P^3$ linklessly. The graph G contains a 0-homologous 3-cycle, since otherwise G contains two disjoint 1-homologous cycles and is linked. Without loss of generality, let $(4, 5, 6)$ be a 0-homologous 3-cycle. Proposition 7 implies that the cycle $(1, 2, 3)$ is 1-homologous as it is disjoint from $(4, 5, 6)$.

Mroczkowski [2003] showed that every cycle can be made, via crossing changes and ambient isotopy, into an unknotted 0-cycle or the 1-homologous cycle g as

explained in the Introduction. Apply crossing changes and ambient isotopy so that the embedding has a projection with vertices as drawn in Figure 3. A priori, the edges between vertices $\{1, 2, 3\}$ and $\{4, 5, 6\}$ may be more complicated than as drawn in the figure.

Vertices $\{1, 2, 3\}$ and the 3-cycle $(1, 2, 3)$ lie on the boundary. In the projection, we label the pair of antipodal identified vertices by $\{v_A, v_B\}$ for $v \in \{1, 2, 3\}$.

Consider the edge E between 1 and 4. Together with the path $(1_B, 4)$ pictured in Figure 3, it forms either a 0-homologous or a 1-homologous cycle. If the cycle formed is 0-homologous, then by Mroczkowski's result, $E \cup (1_B, 4)$ can be made into the unknot by crossing changes, and then deformed so that E is within a small neighborhood of the path $(1_B, 4)$. That is, the cycle does not cross the boundary of D^2. If $E \cup (1_B, 4)$ forms a 1-homologous cycle, then E and the path formed by connecting 4 to 1_A by a straight line segment form a 0-homologous cycle. By similar reasoning, the edge E can be deformed, by crossing changes and ambient isotopy, to be within a small neighborhood of $(1_A, 4)$; that is to say, it does not cross the boundary of D^2. By similar reasoning, all edges between vertices $\{1, 2, 3\}$ and $\{4, 5, 6\}$ may be drawn in the projection onto $\mathbb{R}P^2$ without crossing the boundary.

We now describe how vertices $\{1, 2, 3\}$ connect to vertices $\{4, 5, 6\}$. We use that G does not contain two disjoint 1-homologous cycles or a 0-homologous K_4 by Lemma 2.

Let $v \in \{1, 2, 3\}$. Then v connects to one of $\{4, 5, 6\}$ from v_A and connects to one of $\{4, 5, 6\}$ from v_B; otherwise, G has a 0-homologous K_4 subgraph. Without loss of generality, suppose that $\{2_A\}$ connects to $\{5\}$ and $\{2_B\}$ connects to $\{4\}$ and $\{6\}$.

If $\{1_A\}$ or $\{1_B\}$ connects to both $\{4\}$ and $\{6\}$, then $G[1, 2, 4, 6]$ is a 0-homologous K_4. Thus, without loss of generality, let $\{1_B\}$ connect to $\{4\}$ and $\{1_A\}$ connect to $\{6\}$.

Vertex $\{1_A\}$ connects to $\{5\}$ since otherwise, any arrangement of edges connecting vertex $\{3\}$ to vertices $\{4, 5, 6\}$ induces either two disjoint 1-homologous cycles or a 0-homologous K_4 subgraph, as shown in the table below.

vertices $\{3_A\}$ connects to	vertices $\{3_B\}$ connects to	1-homologous cycles or 0-homologous K_4
$\{4\}$	$\{5\}, \{6\}$	$(1, 3, 6), (2, 4, 5)$
$\{5\}$	$\{4\}, \{6\}$	$G[2, 3, 4, 6]$
$\{6\}$	$\{4\}, \{5\}$	$G[1, 3, 4, 5]$
$\{4\}, \{5\}$	$\{6\}$	$(1, 3, 6), (2, 4, 5)$
$\{4\}, \{6\}$	$\{5\}$	$G[2, 3, 4, 6]$
$\{5\}, \{6\}$	$\{4\}$	$(1, 2, 5), (3, 4, 6)$

Finally, the following table shows that vertex $\{3_A\}$ connects to $\{6\}$ and vertex $\{3_B\}$ connects to 4 and 5. Indeed, all other arrangements lead to either two disjoint 1-homologous cycles or a 0-homologous K_4 subgraph.

vertices $\{3_A\}$ connects to	vertices $\{3_B\}$ connects to	1-homologous cycles or 0-homologous K_4
$\{4\}$, $\{5\}$	$\{6\}$	$(1, 3, 6)$, $(2, 4, 5)$
$\{4\}$, $\{6\}$	$\{5\}$	$G[2, 3, 4, 6]$
$\{5\}$, $\{6\}$	$\{4\}$	$G[1, 3, 5, 6]$
$\{4\}$	$\{5\}$, $\{6\}$	$G[1, 3, 4, 6]$
$\{5\}$	$\{4\}$, $\{6\}$	$G[2, 3, 4, 6]$

Thus, up to crossing changes and ambient isotopy, Figure 3 depicts the only way K_6 may be linklessly embedded in $\mathbb{R}P^3$. □

Introduced by Harary [1953], *signed graphs* are graphs with each edge assigned a $+$ or a $-$ sign, and constitute the final tool in our proof that K_{10} is intrinsically triple-linked in $\mathbb{R}P^3$. An embedding of a graph G into $\mathbb{R}P^3$ induces a signed graph as follows: deform the embedding to that no vertices touch the bounding sphere in the model of $\mathbb{R}P^3$ with $\partial(D^3) \cong S^2$ and so that all intersections of edges with the bounding sphere are transverse. Define an edge of G to be a $^+$edge if the edge intersects the boundary an even number of times and to be a $^-$edge if the edge intersects the boundary an odd number of times. An example is drawn in Figure 4. Note that a cycle with an odd number of $^-$edges is 1-homologous.

Two embeddings G_1 and G_2 of a graph G are *crossing-change equivalent* if G_1 can be obtained from G_2 by crossing changes and ambient isotopy. By Proposition 8, a linkless K_6 embedded in $\mathbb{R}P^3$ is crossing-change equivalent to the embedding drawn in Figure 4. That is, if G is a signed graph isomorphic to K_6 with vertex set $\{1, 2, 3, 4, 5, 6\}$, then G is crossing-change equivalent to a signed graph with $^-$edge set $S = \{(1, 2), (1, 3), (2, 3), (1, 4), (2, 5), (3, 6)\}$ and $^+$edge set $E(G) \backslash S$.

Our next result shows that if G is a graph isomorphic to K_{10}, then G is intrinsically triple-linked in $\mathbb{R}P^3$. We first sketch an outline. Using results of [Flapan et al. 2001a; Bowlin and Foisy 2004], we show that a 3-linkless embedding of G, if such an
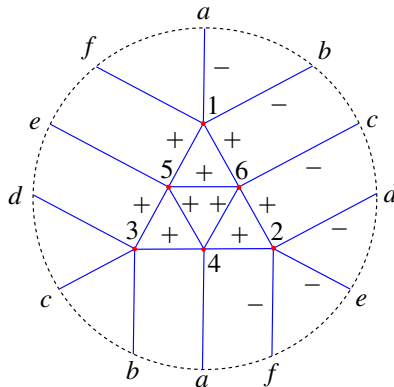
Figure 4. A signed linkless embedding of K_6 in $\mathbb{R}P^3$.

embedding exists, must contain a linkless K_6 subgraph. We prove the remaining four vertices must induce a 0-homologous K_4 subgraph or the embedded graph contains a nonsplit triple link. Finally, we determine the signs of the edges connecting the K_6 subgraph to the K_4 subgraph, eventually determining that any possible sign assignment results in a triple link. Thus, no 3-linkless embedding of G can exist.

Theorem 9. *The graph K_{10} is intrinsically triple-linked in $\mathbb{R}P^3$.*

Proof. Let G be a graph isomorphic to K_{10} with vertex set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Embed G in $\mathbb{R}P^3$ as a signed graph and assume, toward a contradiction, that G is 3-linkless.

If every subgraph of G isomorphic to K_6 is linked, then the proof in [Flapan et al. 2001a] that K_{10} is intrinsically linked in $\mathbb{R}^3$ nearly holds in $\mathbb{R}P^3$. However, at the end of their proof, they use that $K_{3,3,1}$ is intrinsically linked in $\mathbb{R}^3$, but this graph embeds linklessly in $\mathbb{R}P^3$. Bowlin and Foisy [2004] modified the proof in [Flapan et al. 2001a] to only use the fact that K_6 is intrinsically linked in $\mathbb{R}^3$. Thus, in the case that every subgraph of G isomorphic to K_6 is linked, G contains a triple link. So, we may assume that there exists a linkless K_6 subgraph of G. Without loss of generality, suppose that $G[1, 2, 3, 4, 5, 6]$ is linkless. By Proposition 8, the subgraph $G[1, 2, 3, 4, 5, 6]$ has an embedding that is crossing-change equivalent to that drawn in Figure 4. In particular, since crossing changes do not change the homology of cycles, we may assume $(1, 2, 3)$ is 1-homologous.

Claim. *The embedded induced subgraph $G[7, 8, 9, 10]$ is 0-homologous.*

Proof. Suppose $G[7, 8, 9, 10]$ has a 1-homologous cycle. Without loss of generality, suppose $(7, 8, 9)$ is 1-homologous. If $G[4, 5, 6, 10]$ is not 0-homologous, then two of $(4, 5, 10)$, $(4, 6, 10)$, and $(5, 6, 10)$ are 1-homologous by Lemma 3, since we have assumed $(4, 5, 6)$ is 0-homologous. Then $(1, 2, 3)$, $(7, 8, 9)$, and a cycle from $G[4, 5, 6, 10]$ comprise three disjoint 1-homologous cycles, so G is triple-linked. Thus, $G[4, 5, 6, 10]$ is 0-homologous and so $G[1, 2, 4, 5, 6, 10]$ has a pair of linked cycles by Lemma 2. Since $(7, 8, 9)$ is 1-homologous, and $(7, 8, 9)$ is disjoint from all the 1-homologous cycles in the second column of Table 1, Lemma 4 applies and G has a triple link. Thus, $G[7, 8, 9, 10]$ is 0-homologous. $\square$

Since ambient isotopy and crossing changes do not change the homology of cycles, we may modify the embedding of G so that all edges in $G[7, 8, 9, 10]$ are $^+$edges and the edges in $G[1, 2, 3, 4, 5, 6]$ are $^+$edges and $^-$edges as defined in Figure 4. Many of the remaining arguments rely on linked K_6 subgraphs of G and use the argument highlighted in Table 1. In particular, though the K_6 subgraph of the modified embedding may contain a different pair of linked cycles than the original embedding, our argument relies only on the existence of linked cycles, not on the specific pair of linked cycles. Thus, we now consider the signs of the edges connecting $G[1, 2, 3, 4, 5, 6]$ to $G[7, 8, 9, 10]$.

possible linked cycles in $G[1, 2, 4, 5, 6, 10]$	1-homologous cycle that shares an edge with a linked cycle
$(1, 2, 4)$, $(5, 6, 10)$	$(1, 2, 3)$
$(1, 2, 5)$, $(4, 6, 10)$	$(1, 2, 3)$
$(1, 2, 6)$, $(4, 5, 10)$	$(1, 2, 3)$
$(1, 2, 10)$, $(4, 5, 6)$	$(1, 2, 3)$
$(1, 4, 5)$, $(2, 6, 10)$	$(1, 3, 5)$
$(1, 4, 6)$, $(2, 5, 10)$	$(1, 4, 6)$
$(1, 4, 10)$, $(2, 5, 6)$	$(2, 5, 6)$
$(1, 5, 6)$, $(2, 4, 10)$	$(1, 3, 5)$
$(1, 5, 10)$, $(2, 4, 6)$	$(1, 3, 5)$
$(1, 6, 10)$, $(2, 4, 5)$	$(2, 4, 5)$

Table 1

Claim. *If $v \in \{1, 2, 3\}$, then edges from v to $G[7, 8, 9, 10]$ have the same sign.*

Proof. Assume toward a contradiction that the edges from $\{1\}$ to $G[7, 8, 9, 10]$ do not all have the same sign. Without loss of generality, let $(1, 7)$ be a $^+$edge and $(1, 8)$ a $^-$edge. Then $(1, 7, 8)$ is a 1-homologous cycle.

Consider $G[3, 4, 6, 9]$. Since $(3, 4, 6)$ is 1-homologous, $G[3, 4, 6, 9]$ contains another 1-homologous cycle by Lemma 3. If $(3, 4, 9)$ or $(3, 6, 9)$ is 1-homologous then the set $\{(1, 7, 8), (2, 5, 6), (3, 4, 9)\}$ or $\{(1, 7, 8), (2, 4, 5), (3, 6, 9)\}$ contains three disjoint 1-homologous cycles, respectively, and so G is triple-linked. Thus, $(4, 6, 9)$ is the second 1-homologous cycle in $G[3, 4, 6, 9]$.

Since $(2, 3, 4)$ is 1-homologous, the induced subgraph $G[2, 3, 4, 9]$ contains a second 1-homologous cycle by Lemma 3. As shown above, $(3, 4, 9)$ is 0-homologous. If $(2, 4, 9)$ is 1-homologous, then $(1, 7, 8)$, $(2, 4, 9)$, and $(3, 5, 6)$ form three disjoint 1-homologous cycles, so G is triple-linked. So, $(2, 3, 9)$ is the second 1-homologous cycle in $G[2, 3, 4, 9]$.

Similarly, since $(3, 5, 6)$ is 1-homologous, $G[3, 5, 6, 9]$ contains a second 1-homologous cycle by Lemma 3. As shown above, $(3, 6, 9)$ is 0-homologous. Additionally, $(5, 6, 9)$ is 0-homologous; otherwise $(1, 7, 8)$, $(2, 3, 4)$, and $(5, 6, 9)$ form three disjoint 1-homologous cycles and G is triple-linked. Thus, $(3, 5, 9)$ is a 1-homologous cycle.

As $(1, 7, 8)$ and $(4, 6, 9)$ are 1-homologous, $G[2, 3, 5, 10]$ is a 0-homologous K_4, since, otherwise, G contains three disjoint 1-homologous cycles. By Lemma 2, $G[2, 3, 4, 5, 6, 10]$ contains a pair of linked cycles. Since $(1, 7, 8)$ is 1-homologous and disjoint from all of the 1-homologous cycles in the second column of Table 2, Lemma 4 applies and G contains a triple-link, a contradiction.

Thus, $\{1\}$ connects to $G[7, 8, 9, 10]$ via all $^+$edges or all $^-$edges, and similar reasoning applies to vertices $\{2\}$ and $\{3\}$. $\square$

possible linked cycles in $G[2, 3, 4, 5, 6, 10]$	1-homologous cycle that shares an edge with a linked cycle
(2, 3, 4), (5, 6, 10)	(2, 3, 4)
(2, 3, 5), (4, 6, 10)	(4, 6, 9)
(2, 3, 6), (4, 5, 10)	(2, 3, 9)
(2, 3, 10), (4, 5, 6)	(2, 3, 9)
(2, 4, 5), (3, 6, 10)	(2, 4, 5)
(2, 4, 6), (3, 5, 10)	(4, 6, 9)
(2, 4, 10), (3, 5, 6)	(3, 5, 9)
(2, 5, 6), (3, 4, 10)	(2, 5, 6)
(2, 5, 10), (3, 4, 6)	(4, 6, 9)
(2, 6, 10), (3, 4, 5)	(3, 5, 9)

Table 2

A similar argument, using different induced subgraphs, shows the edges between each of the remaining vertices of $G[1, 2, 3, 4, 5, 6]$ and $G[7, 8, 9, 10]$ also have the same sign.

Claim. *If $v \in \{4, 5, 6\}$, then all edges from v to $G[7, 8, 9, 10]$ have the same sign.*

Proof. Towards a contradiction, suppose not all the edges from $\{4\}$ to $G[7, 8, 9, 10]$ have the same sign. Without loss of generality, let $(4, 7)$ be a $^+$edge and $(4, 8)$ be a $^-$edge. Then $(4, 7, 8)$ is a 1-homologous cycle.

Since $(1, 2, 3)$ is a 1-homologous cycle, the subgraph $G[1, 2, 3, 9]$ contains a second 1-homologous cycle by Lemma 3. If $(1, 3, 9)$ or $(1, 2, 9)$ is 1-homologous, then the set $\{(1, 3, 9), (2, 5, 6), (4, 7, 8)\}$ or $\{(1, 2, 9), (3, 5, 6), (4, 7, 8)\}$ contains three disjoint 1-homologous cycles, respectively. So, $(2, 3, 9)$ is the second 1-homologous cycle in $G[1, 2, 3, 9]$.

Since $(2, 3, 9)$ and $(4, 7, 8)$ are 1-homologous, the subgraph $G[1, 5, 6, 10]$ is a 0-homologous K_4; otherwise, G contains three disjoint 1-homologous cycles. By Lemma 2, $G[1, 2, 3, 5, 6, 10]$ contains a pair of linked cycles. Since $(4, 7, 8)$ is 1-homologous and disjoint from all 1-homologous cycles in the second column of Table 3, Lemma 4 applies and G contains a triple link.

Thus, all edges from $\{4\}$ to $G[7, 8, 9, 10]$ have the same sign. A similar argument shows that all edges from vertices $\{5\}$ and $\{6\}$ to $G[7, 8, 9, 10]$ have the same sign. $\square$

The previous two claims show that the edges from each vertex in $G[1, 2, 3, 4, 5, 6]$ to the vertices of $G[7, 8, 9, 10]$ have the same sign. As we have assigned signs to the edges of $G[1, 2, 3, 4, 5, 6]$ and $G[7, 8, 9, 10]$, there remain 2^6 possible embedding classes. We consider all cases. If all edges from vertex $v \in \{1, 2, 3, 4, 5, 6\}$ to $G[7, 8, 9, 10]$ are $^+$edges, we write v_+, and otherwise v_-. For v_x with $x \in \{+, -\}$, we say "the sign of vertex v is x".

possible linked cycles in $G[1, 2, 3, 5, 6, 10]$	1-homologous cycle that shares an edge with a linked cycle
$(1, 2, 3), (5, 6, 10)$	$(1, 2, 3)$
$(1, 2, 5), (3, 6, 10)$	$(2, 5, 9)$
$(1, 2, 6), (3, 5, 10)$	$(1, 2, 6)$
$(1, 2, 10), (3, 5, 6)$	$(3, 5, 6)$
$(1, 3, 5), (2, 6, 10)$	$(1, 3, 5)$
$(1, 3, 6), (2, 5, 10)$	$(2, 5, 9)$
$(1, 3, 10), (2, 5, 6)$	$(2, 5, 9)$
$(1, 5, 6), (2, 3, 10)$	$(2, 3, 9)$
$(1, 5, 10), (2, 3, 6)$	$(2, 3, 9)$
$(1, 6, 10), (2, 3, 5)$	$(2, 3, 9)$

Table 3

Claim. *The two vertices in each of the pairs $\{1, 4\}$, $\{2, 5\}$, and $\{3, 6\}$ have different signs.*

Proof. Suppose toward a contradiction that $\{1\}$ and $\{4\}$ are both $^+$edges. Then $(1, 4, 7)$ is a 1-homologous cycle.

Since both $(2, 5)$ and $(3, 6)$ are $^-$edges, if both pairs of vertices $\{2, 5\}$ and $\{3, 6\}$ share the same sign (e.g., $2_+, 5_+, 3_-, 6_-$), then $(2, 5, 8)$ and $(3, 6, 9)$ are 1-homologous cycles. Thus, the cycles $(1, 4, 7)$, $(2, 5, 8)$, and $(3, 6, 9)$ are disjoint and 1-homologous, so G is triple-linked.

Since both $(2, 6)$ and $(3, 5)$ are $^+$edges, if both pairs of vertices $\{2, 6\}$ and $\{3, 5\}$ have different signs (e.g., $2_+, 6_-, 3_+, 5_-$), then $(2, 6, 8)$ and $(3, 5, 9)$ are 1-homologous cycles. Thus, we know $(1, 4, 7)$, $(2, 6, 8)$, and $(3, 5, 9)$ are disjoint 1-homologous cycles, so G is triple-linked.

The edge $(2, 3)$ is a $^-$edge and $(5, 6)$ is a $^+$edge, so if $\{2\}$ and $\{3\}$ share the same sign and $\{5\}$ and $\{6\}$ have different signs, (e.g., $2_+, 3_+, 5_+, 6_-$), then $(2, 3, 8)$ and $(5, 6, 9)$ are 1-homologous cycles. So, $(1, 4, 7)$, $(2, 3, 8)$, and $(5, 6, 9)$ form disjoint 1-homologous cycles, so G is triple-linked.

If G is embedded with $\{2_-, 3_+, 5_+, 6_-\}$ or $\{2_+, 3_-, 5_-, 6_+\}$, then $(1, 4, 7)$ and $(5, 6, 8)$ are disjoint 1-homologous cycles, so $G[2, 3, 9, 10]$ is a 0-homologous K_4, or G has a triple link. So, by Lemma 2, $G[1, 2, 3, 4, 9, 10]$ has a pair of linked cycles. Since $(5, 6, 8)$ is 1-homologous and is disjoint from all of the 1-homologous cycles in the second column of Table 4, G has a triple link by Lemma 4.

Finally, if G is embedded with one of the remaining configurations,

$$\{2_-, 3_+, 5_+, 6_+\}, \quad \{2_-, 3_+, 5_-, 6_-\}, \quad \{2_+, 3_-, 5_-, 6_-\}, \quad \{2_-, 3_+, 5_+, 6_+\},$$

then one of $\{(2, 5, 6), (3, 5, 6)\}$ must be 1-homologous. Since $G[7, 8, 9, 10]$ is a 0-homologous K_4, the subgraph $G[1, 4, 7, 8, 9, 10]$ contains a pair of linked cycles

possible linked cycles in $G[1, 2, 3, 4, 9, 10]$	1-homologous cycle that shares an edge with a linked cycle
$(1, 2, 3), (4, 9, 10)$	$(1, 2, 3)$
$(1, 2, 4), (3, 9, 10)$	$(1, 4, 7)$
$(1, 2, 9), (3, 4, 10)$	$(1, 2, 7)$
$(1, 2, 10), (3, 4, 9)$	$(1, 2, 7)$
$(1, 3, 4), (2, 9, 10)$	$(1, 4, 7)$
$(1, 3, 9), (2, 4, 10)$	$(1, 3, 7)$
$(1, 3, 10), (2, 4, 9)$	$(1, 3, 7)$
$(1, 4, 9), (2, 3, 10)$	$(1, 4, 7)$
$(1, 4, 10), (2, 3, 9)$	$(1, 4, 7)$
$(1, 9, 10), (2, 3, 4)$	$(2, 3, 4)$

Table 4

by Lemma 2. Both $(2, 5, 6)$ and $(3, 5, 6)$ are disjoint from one 1-homologous cycle in each row of the second column of Table 5. Thus, by Lemma 4, G is triple-linked.

So, in each embedding of G with 1_+ and 4_+, the graph G contains a triple link. A similar argument holds in the case that G is embedded with 1_- and 4_- and for the other vertex pairs $\{2, 5\}$ and $\{3, 6\}$. $\qquad\square$

We now suppose G is embedded with 1_+ and 4_-. By the last claim, the vertices in each of the pairs $\{2, 5\}$ and $\{3, 6\}$ have different signs. So, there are four cases to consider:

$$\{2_+, 3_+, 5_-, 6_-\}, \quad \{2_+, 3_-, 5_-, 6_+\}, \quad \{2_-, 3_+, 5_+, 6_-\}, \quad \{2_-, 3_-, 5_+, 6_+\}.$$

possible linked cycles in $G[1, 4, 7, 8, 9, 10]$	1-homologous cycles that share an edge with a linked cycle
$(1, 4, 7), (8, 9, 10)$	$(1, 4, 7)$
$(1, 4, 8), (7, 9, 10)$	$(1, 4, 8)$
$(1, 4, 9), (7, 8, 10)$	$(1, 4, 9)$
$(1, 4, 10), (7, 8, 9)$	$(1, 4, 10)$
$(1, 7, 8), (4, 9, 10)$	$(1, 2, 7), (1, 3, 7)$
$(1, 7, 9), (4, 8, 10)$	$(1, 2, 7), (1, 3, 7)$
$(1, 7, 10), (4, 8, 9)$	$(1, 2, 7), (1, 3, 7)$
$(1, 8, 9), (4, 7, 10)$	$(1, 2, 8), (1, 3, 8)$
$(1, 8, 10), (4, 7, 9)$	$(1, 2, 8), (1, 3, 8)$
$(1, 9, 10), (4, 7, 8)$	$(1, 2, 9), (1, 3, 9)$

Table 5

possible linked cycles in $G[4, 6, 7, 8, 9, 10]$	1-homologous cycle that shares an edge with a linked cycle
$(4, 6, 7), (8, 9, 10)$	$(4, 6, 7)$
$(4, 6, 8), (7, 9, 10)$	$(4, 6, 8)$
$(4, 6, 9), (7, 8, 10)$	$(4, 6, 9)$
$(4, 6, 10), (7, 8, 9)$	$(4, 6, 10)$
$(4, 7, 8), (6, 9, 10)$	$(5, 6, 9)$
$(4, 7, 9), (6, 8, 10)$	$(5, 6, 8)$
$(4, 7, 10), (6, 8, 9)$	$(5, 6, 8)$
$(4, 8, 9), (6, 7, 10)$	$(5, 6, 7)$
$(4, 8, 10), (6, 7, 9)$	$(5, 6, 7)$
$(4, 9, 10), (6, 7, 8)$	$(5, 6, 7)$

Table 6

First, if the embedding has $\{2_+, 3_+, 5_-, 6_-\}$, then $(1, 6, 7)$, $(2, 4, 9)$, and $(3, 5, 8)$ form three disjoint 1-homologous cycles, so G is triple-linked. Second, suppose the embedding has $\{2_+, 3_-, 5_-, 6_+\}$ or $\{2_-, 3_+, 5_+, 6_-\}$. Then the second column of Table 6 contains 1-homologous cycles. Since $G[7, 8, 9, 10]$ is 0-homologous, $G[4, 6, 7, 8, 9, 10]$ has a pair of linked cycles by Lemma 2. Since $(1, 2, 3)$ is 1-homologous and disjoint from all 1-homologous cycles in the second column of Table 6, Lemma 4 applies and G contains a triple link.

Finally, if the embedding has $\{2_-, 3_-, 5_+, 6_+\}$, then the second column of Table 7 contains 1-homologous cycles. As above, since $(1, 2, 3)$ is 1-homologous and disjoint from all 1-homologous cycles in the second column of Table 7, Lemma 4 applies and G contains a triple link.

possible linked cycles in $G[4, 6, 7, 8, 9, 10]$	1-homologous cycle that shares an edge with a linked cycle
$(4, 6, 7), (8, 9, 10)$	$(4, 6, 7)$
$(4, 6, 8), (7, 9, 10)$	$(4, 6, 8)$
$(4, 6, 9), (7, 8, 10)$	$(4, 6, 9)$
$(4, 6, 10), (7, 8, 9)$	$(4, 6, 10)$
$(4, 7, 8), (6, 9, 10)$	$(4, 5, 7)$
$(4, 7, 9), (6, 8, 10)$	$(4, 5, 7)$
$(4, 7, 10), (6, 8, 9)$	$(4, 5, 7)$
$(4, 8, 9), (6, 7, 10)$	$(4, 5, 8)$
$(4, 8, 10), (6, 7, 9)$	$(4, 5, 8)$
$(4, 9, 10), (6, 7, 8)$	$(4, 5, 9)$

Table 7

The same argument holds if G is embedded with 1_- and 4_+. So, for any assignment of signs to the edges from $\{1\}$ and $\{4\}$ to $G[7, 8, 9, 10]$, G contains a triple link, a contradiction. Thus, every embedding of G into $\mathbb{R}P^3$ contains a triple link, so G is intrinsically triple-linked in $\mathbb{R}P^3$. $\qquad\square$

Flapan et al. [2001a] show that K_9 can be embedded 3-linklessly in $\mathbb{R}^3$, and so K_9 can be embedded 3-linklessly in $\mathbb{R}P^3$ as well. Thus, 10 is the smallest n for which K_n is intrinsically triple-linked in $\mathbb{R}P^3$.

4. Other intrinsically triple-linked graphs in $\mathbb{R}P^3$

In this section, we exhibit other intrinsically triple-linked graphs in $\mathbb{R}P^3$. We show that two graphs shown in [Bowlin and Foisy 2004] to be intrinsically triple-linked in $\mathbb{R}^3$ may be embedded 3-linklessly in $\mathbb{R}P^3$. Moreover, the graphs obtained by taking two disjoint copies of these graphs described in [Bowlin and Foisy 2004] give intrinsically triple-linked graphs in $\mathbb{R}P^3$. We begin by describing a family of intrinsically n-linked graphs in $\mathbb{R}P^3$.

Lemma 10. *If an embedded graph has all 0-homologous cycles, then it is crossing-change equivalent to a spatial embedding.*

Proof. Take a spanning tree in the embedded graph. Since a spanning tree is contractible, it can be deformed so that none of its edges touch the boundary of D^2. Order the edges that do not lie in the spanning tree. Now take the first edge not in the spanning tree. If this edge does not touch the boundary, move on to the next edge. Otherwise, the edge lies in a cycle that, by assumption, is 0-homologous. By Mroczkowski's result, the cycle can be made into an unknot by crossing changes. Since the unknot is 0-homologous, it bounds a disk. Deform the edge by pulling in the disk towards the edges of the cycle that lie in the spanning tree. Thus, the edge can be deformed so that it does not touch the boundary of D^2. Eventually, all of the edges not in the spanning tree can be deformed, if necessary, not to touch the boundary. The resulting embedding is equivalent to a spatial embedding. Thus, the original embedding was crossing-change equivalent to a spatial embedding. $\qquad\square$

Proposition 11. *A graph composed of n disjoint copies of an intrinsically n-linked graph in $\mathbb{R}^3$ is intrinsically n-linked in $\mathbb{R}P^3$. In particular, three disjoint copies of intrinsically triple-linked graphs in $\mathbb{R}^3$ are intrinsically triple-linked in $\mathbb{R}P^3$.*

Proof. Let G be a graph that is intrinsically n-linked in $\mathbb{R}^3$, and let G_i be isomorphic to G for $i = 1, \ldots, n$. Let $\Gamma = \bigsqcup_{i=1}^n G_i$ be the disjoint union of n graphs isomorphic to G. If G_i contains all 0-homologous cycles for some i, then G_i is crossing-change equivalent to a spatial embedding by Lemma 10. Thus, G_i, and hence G, is n-linked in $\mathbb{R}P^3$.
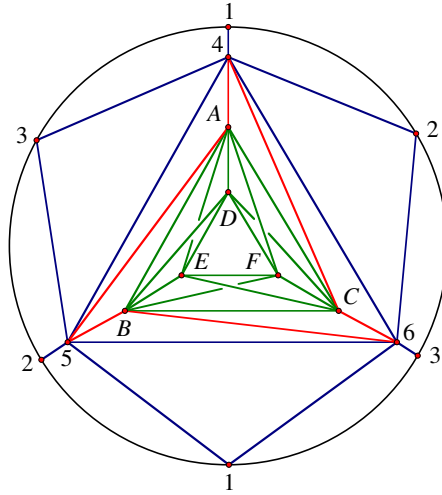
Figure 5. A 3-linkless embedding of K_6 connected to K_6 along a 6-cycle in $\mathbb{R}P^3$.

Otherwise, each G_i contains a 1-homologous cycle. Thus, Γ contains n disjoint 1-homologous cycles, and so contains an n-link. Therefore, Γ is intrinsically n-linked in $\mathbb{R}P^3$. □

The graph K_{10} is an example of a one-component graph that is intrinsically triple-linked in $\mathbb{R}P^3$. We now exhibit two intrinsically triple-linked graphs in $\mathbb{R}P^3$, each comprised of two components. In each case, the components are intrinsically triple-linked in $\mathbb{R}^3$. The question remains whether there exists a minor-minimal intrinsically triple-linked graph of three components in $\mathbb{R}P^3$.

Bowlin and Foisy prove the following graphs are intrinsically linked in $\mathbb{R}^3$.

Theorem 12 [Bowlin and Foisy 2004]. *Let G be a graph containing two disjoint graphs from the Petersen family, G_1 and G_2, as subgraphs. If there are edges between the two subgraphs G_1 and G_2 such that the edges form a 6-cycle with vertices that alternate between G_1 and G_2, then G is minor-minimal intrinsically triple-linked in $\mathbb{R}^3$.*

If G_1 and G_2, as in the theorem, are isomorphic to K_6, this result does not hold in $\mathbb{R}P^3$. A 3-linkless embedding of $G = G_1 \sqcup G_2$ is shown in Figure 5. We now show that the graph obtained from two disjoint copies of G is minor-minimal intrinsically triple-linked in $\mathbb{R}P^3$.

Theorem 13. *Let G_1 be a graph containing two disjoint copies of K_6 with edges between the two K_6 subgraphs that form a 6-cycle with vertices alternating between the two K_6 subgraphs. If G_2 is a graph isomorphic to G_1 and $G = G_1 \sqcup G_2$, then G is minor-minimal intrinsically triple-linked in $\mathbb{R}P^3$.*

Proof. Let $G = G_1 \sqcup G_2$ be as in the theorem, and embed G in $\mathbb{R}P^3$.

If either G_1 or G_2 contain all 0-homologous cycles, then that subgraph is crossing-change equivalent to a spatial embedding by Lemma 10, and hence triple-linked by Theorem 12. Thus, G contains a triple link. So, now suppose that both G_1 and G_2 contain a 1-homologous cycle.

In both G_1 and G_2, any cycle of length greater than 3 can be subdivided by an edge e into a "θ-graph": two cycles of smaller length, disjoint, except for edge e. That is, there exists an edge $e = (v_1, v_i)$ in $G[v_1, \ldots, v_n]$ so that $c = (v_1, \ldots, v_n)$ may be divided into $c_1 \cup c_2 = (v_1, \ldots, v_i) \cup (v_i, \ldots, v_n, v_1)$. If c is 1-homologous, then in any signed embedding of G, the cycle c has an odd number of $^-$edges. So, either c_1 or c_2 has an odd number of $^-$edges, and is thus 1-homologous. By iterating this procedure, we conclude that both G_1 and G_2 contain a 1-homologous 3-cycle.

Let the vertex set of G_1 be given by $\{1, 2, 3, 4, 5, 6, A, B, C, D, E, F\}$ so that $G[1, 2, 3, 4, 5, 6] \cong K_6$ and $G[A, B, C, D, E, F] \cong K_6$ are connected by edges $(4, A)$, $(4, C)$, $(5, A)$, $(5, B)$, $(6, B)$, and $(6, C)$. Up to isomorphism, there are five 3-cycle equivalence classes in G_1. The set

$$S = \{(1, 2, 3), (1, 2, 4), (1, 4, 5), (4, 5, 6), (4, 5, A)\}$$

contains one representative from each 3-cycle equivalence class. So, without loss of generality, we may suppose that S contains a 1-homologous 3-cycle.

If $G[B, C, E, F] \cong K_4$ contains a 1-homologous cycle, then this cycle, the 1-homologous cycle in S and the 1-homologous cycle in G_2 form three disjoint 1-homologous cycles and so G contains a triple link. Now suppose $G[B, C, E, F]$ is 0-homologous, so that $G[A, B, C, D, E, F]$ contains a pair of linked cycles by Lemma 2.

First suppose that the 1-homologous cycle $c_1 \in S$ is not $(4, 5, A)$. By the pigeonhole principle, two vertices in $\{A, B, C\}$ are in one of the components, c_2, of the linked cycles in $G[A, B, C, D, E, F]$. Use the edges of the 6-cycle to join c_2 to c_1 along disjoint paths. By Lemma 5, G contains a triple link.

Now suppose that the 1-homologous cycle in S is $(4, 5, A)$. If there is a 1-homologous cycle in $G[1, 2, 3, 6]$ then this cycle will link with $(4, 5, A)$ and the 1-homologous cycle in G_2, so G contains a triple link. Else, $G[1, 2, 3, 4, 5, 6]$ has a pair of linked cycles by Lemma 2. By the pigeonhole principle, at least two vertices in the set $\{4, 5, 6\}$ are in a linked cycle, c_3, within $G[1, 2, 3, 4, 5, 6]$. Similarly, at least two vertices of $\{A, B, C\}$ are in a linked cycle, c_4, within $G[A, B, C, D, E, F]$. As a result of the 6-cycle connecting these two copies of K_6, there are two disjoint edges between c_3 and c_4. By Lemma 5, G is triple-linked.

To see G is minor-minimal with respect to intrinsic triple-linking in $\mathbb{R}P^3$, embed G so that G_1 is embedded as in the drawing in Figure 5 and G_2 is contained in a sphere that lies in the complement of G_1. Therefore, G_1 does not have any triple
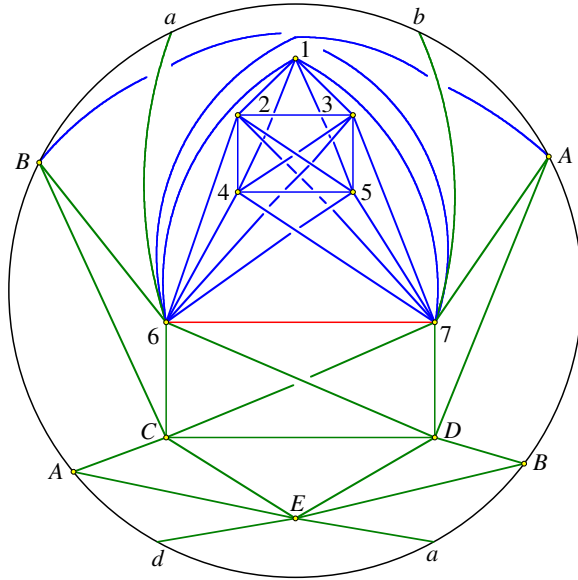
Figure 6. A 3-linkless embedding of K_7 connected to K_7 along an edge in $\mathbb{R}P^3$.

links and no cycle in G_1 is linked with a cycle in G_2. Without loss of generality, if we delete an edge, contract an edge or delete any vertex on G_2, it will have an affine linkless embedding. Thus, we can re-embed G_2 within the sphere in each case. Therefore, G is minor-minimal for intrinsic triple-linking.   □

Bowlin and Foisy prove the following graph is intrinsically triple-linked in $\mathbb{R}^3$.

Theorem 14 [Bowlin and Foisy 2004]. *Let G be a graph formed by identifying an edge of K_7 with an edge from another copy of K_7. Then G is intrinsically triple-linked in $\mathbb{R}^3$.*

The graph G defined in Theorem 14 may be embedded 3-linklessly in $\mathbb{R}P^3$, as drawn in Figure 6. As in the previous result, the graph consisting of two disjoint copies of this graph is intrinsically linked in $\mathbb{R}P^3$.

Theorem 15. *Let G_1 be a graph formed by identifying an edge of K_7 with an edge from another copy of K_7. If G_2 is isomorphic to G_1 and $G = G_1 \sqcup G_2$ is the disjoint union of G_1 and G_2, then G is intrinsically linked in $\mathbb{R}P^3$.*

Proof. Let $G = G_1 \sqcup G_2$ be as above, and embed G in $\mathbb{R}P^3$. If either G_1 or G_2 contains all 0-homologous cycles, then that subgraph is crossing-change equivalent to a spatial embedding by Lemma 10, and hence triple-linked by Theorem 14. Thus, in this case, G has a triple link. Now suppose that both G_1 and G_2 contain a 1-homologous cycle.

Let the vertex set of G_1 be given by $\{1, 2, 3, 4, 5, 6, 7, A, B, C, D, E\}$ so that $G[1, 2, 3, 4, 5, 6, 7]$ and $G[6, 7, A, B, C, D, E]$ are isomorphic to K_7 and share edge $(6, 7)$. Up to isomorphism, there are three 3-cycle equivalence classes in G_1. The set $S = \{(1, 2, 3), (1, 2, 7), (1, 6, 7)\}$ contains one representative from each 3-cycle equivalence class. By the same argument given in Theorem 13, we may assume that S contains a 1-homologous cycle, c_1.

If $G[A, B, C, D]$ contains a 1-homologous cycle, then this cycle, c_1, and the 1-homologous cycle in G_2 form three disjoint 1-homologous cycles, so G contains a triple link. Otherwise, $G[A, B, C, D, E, 6]$ contains a pair of linked cycles by Lemma 2. Following the proof in Theorem 13, connect the linked cycle containing vertex $\{6\}$ to c_1 via two disjoint paths. By Lemma 5, G contains a triple link. □

The minor-minimality of the graph formed by identifying an edge of K_7 with an edge from another copy of K_7 with respect to intrinsic triple-linking is unknown in $\mathbb{R}^3$. If true, then the graph G defined in Theorem 15 is also minor-minimal with respect to intrinsic triple-linking; a similar argument to that in Theorem 13 holds in this case as well.

We also remark that the graph $G(n)$ as defined in [Flapan et al. 2001b] is a one-component minor-minimal intrinsically $(n+1)$-linked graph in $\mathbb{R}P^3$. The arguments given in [Flapan et al. 2001b] hold in $\mathbb{R}P^3$ since $K_{4,4} - \{e\}$, where e is an edge, is intrinsically linked in both $\mathbb{R}^3$ and $\mathbb{R}P^3$.

5. Graphs with linking number at least 1 in $\mathbb{R}P^3$

In $\mathbb{R}P^3$, there are intrinsically linked graphs for which there exists an embedding in which every pair of disjoint cycles has linking number less than 1, as a pair of linked cycles may have only one crossing. Work has been done in $\mathbb{R}^3$ [Flapan 2002] to find graphs containing disjoint cycles with large linking number in every spatial embedding. Using the fact that K_{10} is triple-linked in $\mathbb{R}^3$, Flapan [2002] showed that every spatial embedding of K_{10} contains a two-component link $L \cup J$ such that, for some orientation, $\mathrm{lk}(L, J) \geq 2$. A similar argument using Theorem 9 yields the following proposition.

Proposition 16. *Every projective embedding of K_{10} contains a two-component link $L \cup J$ such that, for some orientation, $\mathrm{lk}(L, J) \geq 1$.*

It remains an open question to determine whether 10 is the smallest number for which this property holds. At this point, we know the smallest n is such that $7 < n \leq 10$.

References

[Bowlin and Foisy 2004] G. Bowlin and J. Foisy, "Some new intrinsically 3-linked graphs", *J. Knot Theory Ramifications* **13**:8 (2004), 1021–1027. MR Zbl

[Bustamante et al. 2009] J. Bustamante, J. Federman, J. Foisy, K. Kozai, K. Matthews, K. McNamara, E. Stark, and K. Trickey, "Intrinsically linked graphs in projective space", *Algebr. Geom. Topol.* **9**:3 (2009), 1255–1274. MR Zbl

[Conway and Gordon 1983] J. H. Conway and C. M. Gordon, "Knots and links in spatial graphs", *J. Graph Theory* **7**:4 (1983), 445–453. MR Zbl

[Drobotukhina 1990] Y. V. Drobotukhina, "An analogue of the Jones polynomial for links in $\mathbf{R}P^3$ and a generalization of the Kauffman–Murasugi theorem", *Algebra i Analiz* **2**:3 (1990), 171–191. In Russian; translated in *Leningrad Math. J.* **2**:3 (1991), 613–630. MR Zbl

[Flapan 2002] E. Flapan, "Intrinsic knotting and linking of complete graphs", *Algebr. Geom. Topol.* **2** (2002), 371–380. MR Zbl

[Flapan et al. 2001a] E. Flapan, R. Naimi, and J. Pommersheim, "Intrinsically triple linked complete graphs", *Topology Appl.* **115**:2 (2001), 239–246. MR Zbl

[Flapan et al. 2001b] E. Flapan, J. Pommersheim, J. Foisy, and R. Naimi, "Intrinsically *n*-linked graphs", *J. Knot Theory Ramifications* **10**:8 (2001), 1143–1154. MR Zbl

[Flapan et al. 2006] E. Flapan, H. Howards, D. Lawrence, and B. Mellor, "Intrinsic linking and knotting of graphs in arbitrary 3-manifolds", *Algebr. Geom. Topol.* **6** (2006), 1025–1035. MR Zbl

[Harary 1953] F. Harary, "On the notion of balance of a signed graph", *Michigan Math. J.* **2** (1953), 143–146. MR Zbl

[Manturov 2004] V. Manturov, *Knot theory*, Chapman & Hall/CRC, Boca Raton, FL, 2004. MR Zbl

[Mroczkowski 2003] M. Mroczkowski, "Diagrammatic unknotting of knots and links in the projective space", *J. Knot Theory Ramifications* **12**:5 (2003), 637–651. MR Zbl

[Nešetřil and Thomas 1985] J. Nešetřil and R. Thomas, "A note on spatial representation of graphs", *Comment. Math. Univ. Carolin.* **26**:4 (1985), 655–659. MR Zbl

[Robertson and Seymour 2004] N. Robertson and P. D. Seymour, "Graph minors, XX: Wagner's conjecture", *J. Combin. Theory Ser. B* **92**:2 (2004), 325–357. MR Zbl

[Robertson et al. 1995] N. Robertson, P. Seymour, and R. Thomas, "Sachs' linkless embedding conjecture", *J. Combin. Theory Ser. B* **64**:2 (1995), 185–227. MR Zbl

[Sachs 1983] H. Sachs, "On a spatial analogue of Kuratowski's theorem on planar graphs—an open problem", pp. 230–241 in *Graph theory* (Lagow, 1981), edited by M. Borowiecki et al., Lecture Notes in Math. **1018**, Springer, Berlin, 1983. MR Zbl

[Sachs 1984] H. Sachs, "On spatial representations of finite graphs", pp. 649–662 in *Finite and infinite sets, II* (Eger, 1981), edited by A. Hajnal et al., Colloq. Math. Soc. János Bolyai **37**, North-Holland, Amsterdam, 1984. MR Zbl

federmanjs@gmail.com   *Department of Mathematics, SUNY Potsdam, Potsdam, NY 13676, United States*

foisyjs@potsdam.edu   *Department of Mathematics, SUNY Potsdam, Potsdam, NY 13676, United States*

mcnam2km@jmu.edu   *Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA 22807, United States*

emily.stark@tufts.edu   *Department of Mathematics, Pomona College, Claremont, CA 91711, United States*

msp

A modified wavelet method
for identifying transient features in time signals
with applications to bean beetle maturation

David McMorris, Paul Pearson and Brian Yurk

(Communicated by Kenneth S. Berenhaut)

We develop an averaging method, based on a modified Haar wavelet technique, for identifying when transient features occur in a time signal. We call this method the seaweed method and use it to identify different stages of bean beetle embryo maturation. We use randomized simulations to evaluate the seaweed method for accuracy and precision at different levels of noise. Our results support the efficacy of the seaweed method as a means for analyzing time-lapse photographs of bean beetle embryos and a wide variety of other time signals.

1. Introduction

Wavelets are commonly employed in signal processing for their usefulness in detecting sudden changes in a signal. In this paper we introduce a new method based on a modification of a Haar wavelet transform. This method, which we call the "seaweed method," is designed to identify points signifying the beginning of a gradual change in the signal, indicating a change in the sign of the slope on a range of time scales. We call such points "transition points," as they often indicate important transitions in the underlying signal. We show this method is effective at identifying transitions in a noisy signal of pixel brightness values coming from time-lapse photography. We believe this method will prove capable of analyzing many kinds of noisy time signals, including audio signals, stock market data, and electrical signals from sensors. Spatial signals, such as population density of a given organism along a transect, may also benefit from analysis by the seaweed method.

The seaweed method was motivated by the need to identify biological markers in developing bean beetle (*Callosobruchus maculatus*) embryos, for the purpose of studying factors that influence insect development time. Ultimately, these studies should yield insights into how climate change will impact insect phenology and,
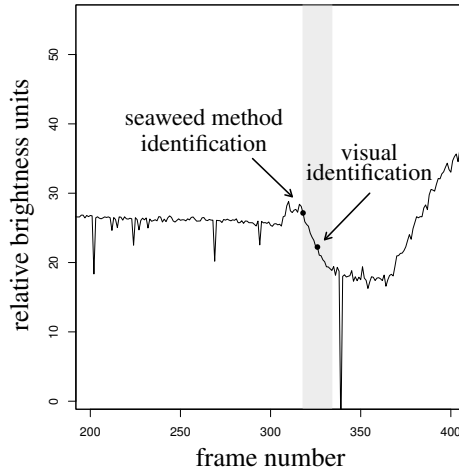
Figure 1. Graph of relative brightness of a particular bean beetle egg over time between frames 200 and 400. Frames are 20 minutes apart. The shaded region indicates the darkening of the embryo's head capsule, and arrows indicate the differences in identification of this marker by eye versus using the seaweed method. Also note that in this signal, the large spikes down are the spurious results of beetles crossing the calibration region on the bean.

hence, population dynamics (see, e.g., [Logan and Powell 2001; Logan et al. 1976; Yurk and Powell 2009]).

Bean beetles are agricultural pests native to Africa and Asia that infest legume crops both in fields and storehouses [Beck and Blumer 2011]. These insects lay their eggs on the surfaces of beans. Following the completion of embryonic development, the larvae burrow into the bean, where they feed and complete development before emerging as adults. In addition to being an important agricultural pest, bean beetles are studied as model organisms, as they are readily obtainable and easy to work with in a laboratory.

In order to better understand the timing of bean beetle embryonic development, we use biological markers to distinguish between different stages of maturation. The marker we use to signify the completion of embryonic development is the formation and darkening of the larval head capsule, which appears as a darkening in the eggs. Although this spot is visible to the naked eye, its formation is so gradual that visual identification of its initial darkening is impossible. Figure 14 shows a sequence of digital microscope images which demonstrates this darkening process.

Because of the difficulties posed by visually identifying the beginning of head-capsule darkening, we used digital time-lapse microscopy to record changes in color and brightness of the eggs at 20-minute intervals. Note that, unlike the images

in Figure 14, these images include many different eggs rather than a single egg. For each egg in the images, we focused on a small, fixed rectangular region of the images containing the egg. We used these images to investigate embryonic development time, and developed a method to pinpoint the precise time when maturation had finished. In each image, the brightness level of an egg, relative to a calibration region on the bean, can be extracted as the difference of two average grayscale values from different regions of the image. In the same way, we can extract a signal consisting of the relative brightness of a specific egg from a sequence of images. A portion of one such signal is displayed in Figure 1. Note that these signals are scaled and shifted for consistency, so the resulting quantities would be dimensionless. Based on our visual observations, we know that the embryo's head capsule darkens along the decline in the shaded region of Figure 1. Whereas a purely visual analysis of the eggs would identify the head-capsule darkening somewhere in the middle of this shaded region, the extraction of this signal and application of the seaweed method make it possible to accurately locate the beginning of this region. This allows for more precise measurement of embryonic development time. In general, the shape of the signal seen in Figure 1 is common to all the eggs we analyze.

The seaweed method is a graphical method based on a modification of the Haar wavelet algorithm. Herein we present the seaweed method, as well as statistical tests to validate it.

2. Background

2A. *Wavelets.* Discrete and continuous wavelet transforms have been used extensively over the last few decades to analyze time signals, as discussed in [Aboufadel and Schlicker 1999; Bénéteau and Van Fleet 2011; Mulcahy 1996]. They provide a way to express a time signal as a sum of component waves. Each component wave has a fixed frequency (or, equivalently, a fixed period) and is a sum of wavelets obtained from a "mother" wavelet by time shifting, time dilation, and amplitude scaling. These wavelet transforms measure changes in amplitude of all component waves over time and frequency simultaneously, thereby allowing signal features to be identified. Signal features include changes in amplitude over a range of time that identifies when an event occurred, changes in amplitude over a range of frequencies that identify the frequency signature of an event, or a combination of the two. This paper will focus on step detection (or change detection) in a time signal using a variation of the Haar wavelet transform.

The discrete wavelet transform separates a signal into a direct sum of wavelets by requiring that amplitude measurements for component waves occur over rectangles of equal area that tile the time-frequency plane but do not overlap, as shown in Figure 2. The width of the rectangles in the top row of Figure 2 is determined by the sampling rate of the signal.
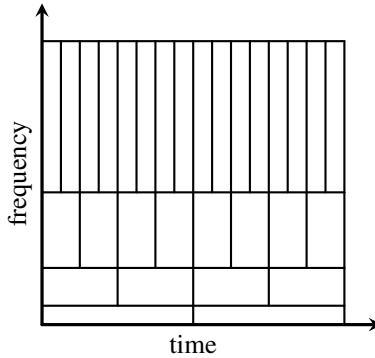
Figure 2. An example of the tiling of the time-frequency plane used for the discrete wavelet transform (DWT). The DWT expresses the signal as a sum of wavelets with one wavelet for each rectangle.

In contrast, the continuous wavelet transform performs more amplitude calculations by increasing the number of rectangular tiles in the time-frequency plane and allowing them to overlap. This overlap in the continuous wavelet transform allows for amplitude measurements in the time-frequency plane to vary continuously both in the time direction and in the frequency direction.

The Haar wavelet transform will be used as the basis for signal processing by wavelets in this paper. The prototype Haar wavelet and Haar scaling function are defined as follows. The Haar "mother" wavelet is the function $\psi : \mathbb{R} \to \mathbb{R}$ defined by

$$\psi(t) = \begin{cases} -1 & \text{if } 0 \leq t < \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-1}$$

Contrary to standard convention, we have chosen $\psi(t)$ to be an increasing, rather than decreasing, step function over $0 \leq t < 1$. This choice will make interpreting amplitude coefficients for Haar wavelets easier since positive coefficients will indicate an increasing function and negative coefficients a decreasing function. The Haar scaling function is the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-2}$$

2B. *Averaging and differencing.* In [Bénéteau and Van Fleet 2011; Mulcahy 1996], the discrete Haar wavelet transformation is described as a process of averaging and "differencing" as follows. Suppose a discrete, real-valued time signal is sampled at a constant rate, that Δt is the time between consecutive signal measurements, and that s_n denotes the n-th signal value ($n \geq 1$). The process of averaging and differencing starts by setting $\alpha_{0,n} = s_n$. Then, for each $j = 1, 2, 3, \ldots$ and each

averages										differences								
$j \backslash n$	1	2	3	4	5	6	7	8		$j \backslash n$	1	2	3	4	5	6	7	8
0	8	20	32	36	32	20	8	4		0								
1		14		34		26		6		1		6		2		-6		-2
2				24				16		2				10				-10
3								20		3								-4

Table 1. Averages and differences arrays.

$n = 2^j, 2 \cdot 2^j, 3 \cdot 2^j, \ldots$, averages $\alpha_{j,n}$ and "differences" $\delta_{j,n}$ are defined by

$$\alpha_{j,n} = \frac{\alpha_{j-1,n-2^{j-1}} + \alpha_{j-1,n}}{2}, \quad \delta_{j,n} = \frac{\alpha_{j-1,n} - \alpha_{j-1,n-2^{j-1}}}{2}. \qquad (2\text{-}3)$$

Each $\alpha_{j,n}$ is the average of 2^j signal samples, while each $\delta_{j,n}$ is half of the difference between two average values of the signal, where each average is of 2^{j-1} signal values. Thus, each $\delta_{j,n}$ measures changes in the average value of the signal. The averages $\alpha_{j,n}$ and differences $\delta_{j,n}$ are amplitude coefficients for time-shifted and time-dilated Haar scaling functions $\phi_{j,n}(t)$ and Haar wavelets $\psi_{j,n}(t)$, defined as

$$\phi_{j,n}(t) = \phi\left(\frac{t - (n - 2^j + 1)\Delta t}{2^j \Delta t}\right), \quad \psi_{j,n}(t) = \psi\left(\frac{t - (n - 2^j + 1)\Delta t}{2^j \Delta t}\right). \quad (2\text{-}4)$$

The time window over which $\phi_{j,n}(t)$ and $\psi_{j,n}(t)$ are nonzero has length $2^j \Delta t$.

Example 2.1. Suppose $s(t)$ is a time signal with amplitude measurements 8, 20, 32, 36, 32, 20, 8, 4 sampled uniformly over the time interval $0 \le t \le 8$, so that $\Delta t = 1$. By (2-3), the averages and differences for this signal are given by the values in Table 1.

The signal can be written as a sum of Haar scaling functions and Haar wavelets over disjoint time intervals of length 2 by using the row $j = 1$ averages and differences as coefficients:

$$s(t) = \sum_{n=2,4,6,8} \alpha_{1,n}\phi_{1,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n}\psi_{1,n}(t). \qquad (2\text{-}5)$$

Graphs of the signal, the Haar scaling functions, and the Haar wavelets from (2-5) are shown in the rows $j = 0, 1$ of Figure 3. Just as the signal was split into a sum of Haar scaling functions and Haar wavelets, the average values in (2-5) can be treated as a time signal and split into a sum of Haar scaling functions and Haar wavelets:

$$s(t) = \sum_{n=4,8} \alpha_{2,n}\phi_{2,n}(t) + \sum_{n=4,8} \delta_{2,n}\psi_{2,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n}\psi_{1,n}(t). \qquad (2\text{-}6)$$
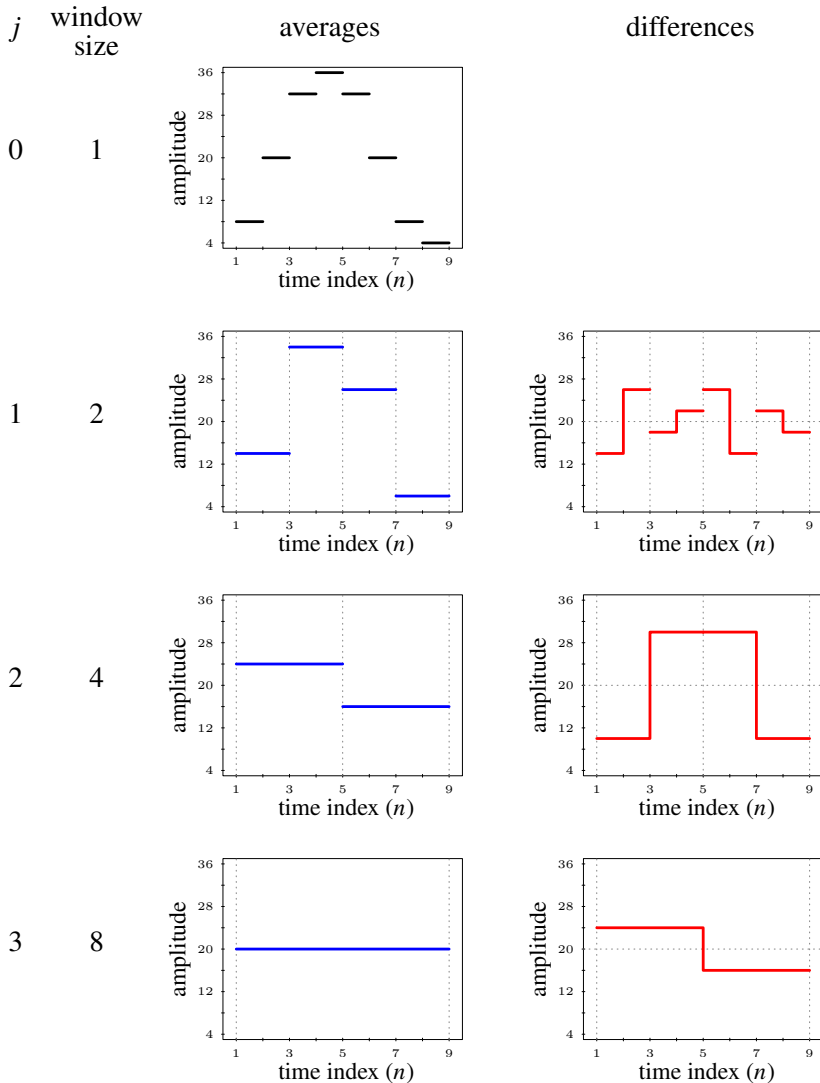
Figure 3. Plot of the signal (row $j = 0$), Haar scaling functions or average values of the signal (left graphs in rows $j = 1, 2, 3$), and Haar wavelets or component waves (right graphs in rows $j = 1, 2, 3$). The amplitudes of the Haar scaling functions and the Haar wavelets come from the coefficients in the averages array and differences array, respectively.

Equation (2-6) shows that the signal is a sum of Haar scaling functions and Haar wavelets over disjoint time intervals of length 4 plus Haar wavelets over disjoint time intervals of length 2, as shown in the rows $j = 1, 2$ of Figure 3. Similarly, the average
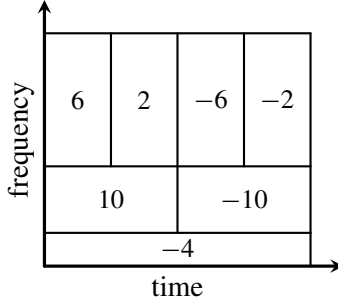
Figure 4. A plot of the Haar wavelet amplitude coefficients $\delta_{j,n}$ in the time-frequency plane can be used to identify features in the signal. For instance, the amplitudes change from positive values (2, 10) to negative values (-6, -10, -4) at time $t = 4$, indicating that the signal changes from increasing to decreasing across multiple frequencies at the same time. Note: the amplitude coefficients in the plot are horizontally centered in each time window (rectangle) to facilitate identifying the time at which a signal feature occurs.

values in (2-6) can be treated as a time signal and split into a sum of Haar scaling functions and Haar wavelets as in (2-7) and in the rows $j = 1, 2, 3$ of Figure 3:

$$s(t) = \sum_{n=8} \alpha_{3,n} \phi_{3,n}(t) + \sum_{n=8} \delta_{3,n} \psi_{3,n}(t) + \sum_{n=4,8} \delta_{2,n} \psi_{2,n}(t) + \sum_{n=2,4,6,8} \delta_{1,n} \psi_{1,n}(t).$$

$$(2\text{-}7)$$

Since each difference $\delta_{j,n}$ represents the amplitude of a Haar wavelet $\psi_{j,n}(t)$ that has frequency $1/(2^j \Delta t)$, the differences $\delta_{j,n}$ can be plotted in the time-frequency plane as a means to identify signal features, as shown in Figure 4.

2C. *Redefining the differences.* The Haar wavelet coefficients $\delta_{j,n}$ are defined recursively in (2-3), but they can be defined explicitly in terms of the signal values:

$$\delta_{j,n} = \frac{1}{2^j} \left(\sum_{i=0}^{2^{j-1}-1} s_{n-i} - \sum_{i=2^{j-1}}^{2^j-1} s_{n-i} \right). \tag{2-8}$$

By writing $1/2^j$ as $\frac{1}{2}(1/2^{j-1})$ and distributing the factor $1/2^{j-1}$ to each term in the sum in (2-8), we see that (2-8) succinctly states that $\delta_{j,n}$ is half of the difference between the average of the 2^{j-1} signal values between s_{n-2^j+1} and $s_{n-2^{j-1}}$ and the average of the 2^{j-1} signal values between $s_{n-2^{j-1}+1}$ and s_n. In other words, $\delta_{j,n}$ is one half of the difference between two average values of the signal over two adjacent time intervals. Each $\delta_{j,n}$ is the amplitude of a Haar wavelet $\psi_{j,n}(t)$ of wavelength $2^j \Delta t$ and frequency $1/(2^j \Delta t)$.

3. Methods

3A. *Modified algorithm.* We now modify the traditional Haar wavelet algorithm by adding more rows to the differences array thereby making it dense, rather than sparse. Adding more rows means having a higher resolution in the frequency domain, while making it dense means using a sliding time window. A sliding window can provide a better resolution in time and is accomplished by allowing n to be any integer, not just a multiple of a power of 2. In this regard, our modified algorithm provides a discrete approximation to the continuous Haar wavelet transform. We define $d_{j,n}$ to be the difference

$$d_{j,n} = \frac{1}{j}\left(\sum_{i=0}^{j-1} s_{n-i} - \sum_{i=j}^{2j-1} s_{n-i} \right). \tag{3-1}$$

In this definition, $2j$ is the size of the window being considered, and with each increase of n, this window is shifted to the right by one time step. Equation (3-1) states that each $d_{j,n}$ is the difference between the average of the j signal values between s_{n-j+1} and s_n and the average of the j signal values between s_{n-2j+1} and s_{n-j}.

There are several benefits of this modified algorithm. This algorithm examines differences in the average values of the signal, making it is easy to describe conceptually. Due to the addition of a sliding window, this method has a higher resolution for detecting signal features than the discrete Haar wavelet transform, as presented in Section 2B. One drawback is that this method is specific to the Haar wavelet and does not immediately generalize to other types of wavelets. Another drawback is that it is more computationally intensive than the discrete Haar wavelet transform, and may even be more computationally intensive than a continuous Haar wavelet transform that has been optimized for speed using the fast Fourier transform.

3B. *Related work.* The modified Haar wavelet algorithm in (3-1) is perhaps best thought of as a discrete approximation to the continuous Haar wavelet transform. It also has some similarities with wavelet frame theory [Christensen 2001; Daubechies et al. 2003; Teolis 1998] and overcomplete discrete wavelet transforms (OCDWT) [Auscher 1992; Bayram and Selesnick 2009; Selesnick 2011; Teolis 1998].

Using Haar wavelet frames, a signal can be represented as a linear combination of wavelets from a frame, which consists of a set of wavelets that span the function space (such as $L^1([a, b])$) but are not necessarily linearly independent, which means the coefficients in the linear combination are not necessarily uniquely determined. Using the modified Haar wavelet algorithm in (3-1), however, the signal is represented uniquely relative to a Haar wavelet basis *within each time window of length* $2j\,\Delta t$, and time windows are allowed to slide (and overlap) by

changing the value of n. A frame could be formed by taking the union of all wavelet bases over all time windows. However, in this study we choose to interpret the wavelet coefficients provided by (3-1) relative to individual time windows that slide through time, rather than relative to the union of all time windows.

The OCDWT and the modified Haar wavelet algorithm in (3-1) are both characterized by denser time-frequency lattices arising from increasing the resolution in the frequency domain, as can be seen in Figure 2 in [Bayram and Selesnick 2009]. The OCDWT has an inherent degree of noise robustness due to redundancy in its representation [Teolis 1998, p. 134], and the modified Haar wavelet algorithm also has this noise robustness due to redundancy. The modified Haar wavelet algorithm has a sliding time window that the OCDWT does not, which means that its time-frequency lattice is more dense than the OCDWT because its time windows slide and are allowed to overlap.

3C. *Identifying transition points.* At this point, we employ the modified algorithm introduced in Section 3A and look for transition points in the signal by locating sign changes between $d_{j,n-1}$ and $d_{j,n}$. When the sign function (the function which takes a number and returns its sign as ± 1 or 0) of the product,

$$p_{j,n} = \text{sgn}(d_{j,n-1} \cdot d_{j,n}) \qquad (3\text{-}2)$$

is nonpositive, either a sign change has occurred or the signal is constant (which in practice is a rare occurrence). A sign change indicates that the corresponding Haar wavelets $d_{j,n-1}\psi_{j,n-1}(t)$ and $d_{j,n}\psi_{j,n}(t)$ changed from either constant or increasing to decreasing or that they changed from either constant or decreasing to increasing at time index n. This process of identifying transitions occurs across multiple time scales as the size of the sliding window increases.

When a transition point is identified, we plot a point at $(n, 2j)$. The type of point we plot depends on the kind of transition that occurs. If the transition signifies an increase in the signal, a closed circle is plotted. If the transition signifies a decrease, an open circle is used. Finally, if the transition signifies that the signal is constant, an open triangle is used. We illustrate this process with the following example.

Example 3.1. Consider the time signal $s(t)$ defined in Example 2.1. The new difference array and corresponding array of product signs for this signal are displayed in Table 2.

Based on the array of products in Table 2, it is clear that transition points have been identified at $n = 5, 6, 7$ using windows of size $2j = 2, 4, 6$ respectively. Figure 5 shows the signal as a piecewise-linear representation as well as the discrete representation with all identified transition points overlaid. The strand of points in Figure 5 identifies the region where the transition occurs in the signal. We call these strands of points "seaweed" due to their visual appearance for more complicated

differences

$j \backslash n$	1	2	3	4	5	6	7	8
1		12	12	4	−4	−12	−12	−4
2				20	8	−8	−20	−20
3						$\frac{28}{3}$	$-\frac{28}{3}$	$-\frac{68}{3}$
4								−8

products

$j \backslash n$	1	2	3	4	5	6	7	8
1				+	+	−	+	+	+
2						+	−	+	+
3								−	+
4								

Table 2. Array of differences between averages on each window
($2j = 2, 4, 6, 8$) as they slide across the time domain, and cor-
responding product sign array. A negative sign indicates that a
transition has occurred.

signals. The presence of multiple points in this strand reveals that this transition is
apparent across multiple scales, and that it is seen with the highest resolution at the
lowest point in the strand. For this reason, the transition point we are looking for is
the lowest point, $(n, 2j) = (5, 2)$. In Section 3D we will introduce guidelines for
identifying transition points in more general cases involving noisier signals.

Figure 6 depicts a more detailed look at the way in which the sign changes are
identified and marked using the algorithm. The first graph in each row shows the last
difference for a certain window that is positive; the second shows the first difference
at that scale that is negative. As the window shifts from a positive difference to
a negative difference, a point is plotted to indicate a transition has occurred. An
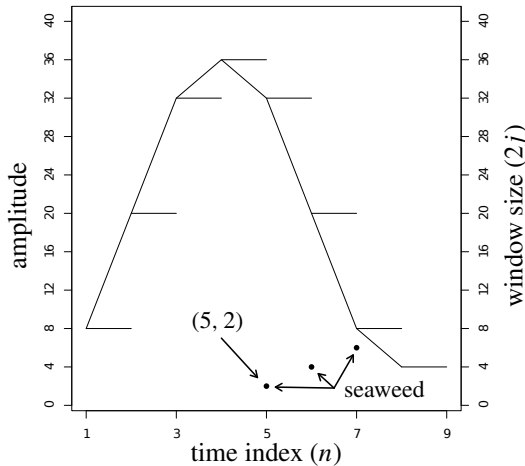
Figure 5. Plot of piecewise-linear representation of the signal used
in Examples 2.1 and 3.1 as well as the discrete representation with
transition points at all time scales marked.
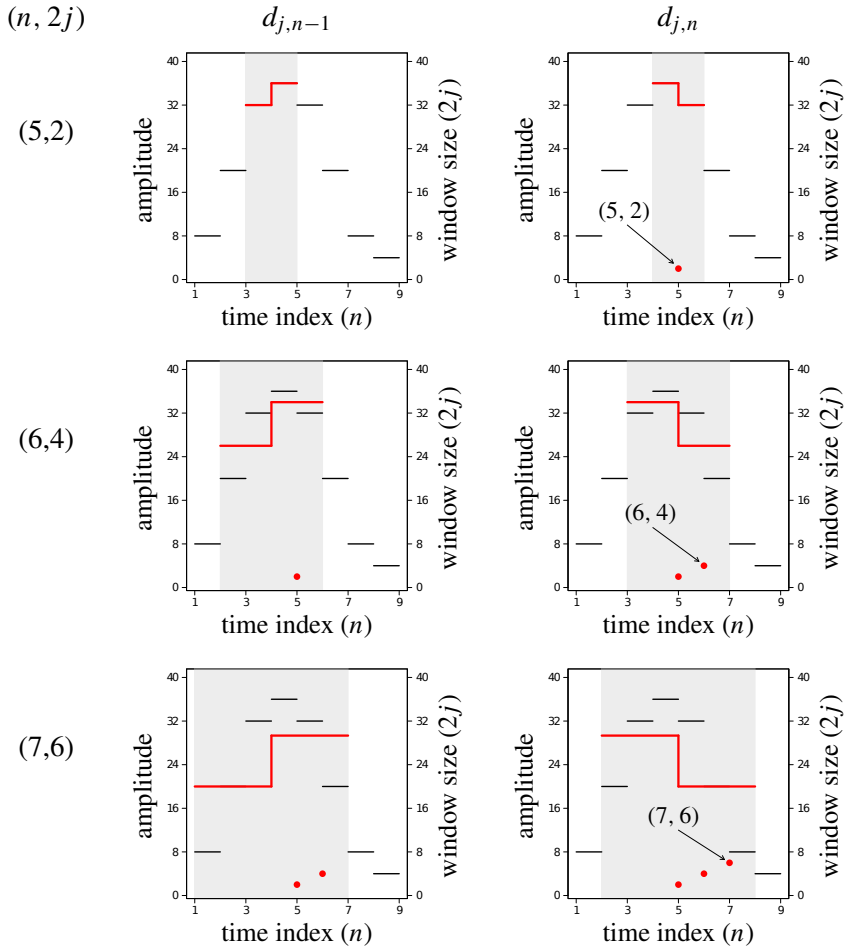
Figure 6. Plots of the signal used in Examples 2.1 and 3.1 depicting the differences in averages on three time scales and the transition points identified as a result of a sign change. The column $d_{j,n-1}$ shows the last positive difference in averages, and the column $d_{j,n}$ shows the first negative difference in averages.

implementation of the seaweed algorithm to the signal in Example 3.1 is included as seaweed_code.R in the online supplement, which can be obtained from this article's publication page.

3D. *Graphical interpretation.* To identify an important transition in the signal with the graphical output, we locate a long branch of seaweed and follow it down until it branches or becomes too hard to distinguish from other, generally shorter, seaweed branches. This happens when the seaweed is being strongly affected by
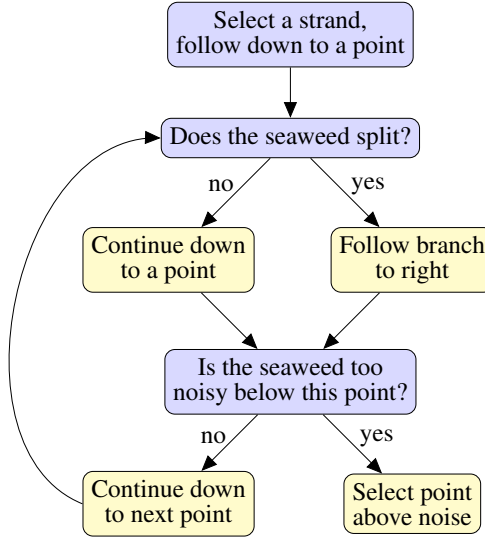
Figure 7. The decision process in the seaweed analysis.

small-scale high-frequency noise, and is no longer being strongly influenced by the global trend. In such a case we see that the seaweed itself becomes too noisy to be reliable at such a small scale. If the branch does split, we follow the general trend, typically on the right side. This is because the algorithm selects the right endpoints of a window, and so selecting the left branch may be departing from a particular strand of seaweed. The time index of the last distinguishable seaweed point on that branch is identified as the time index of a transition point. A decision tree representation of this process is seen in Figure 7.

Next, we examine three additional example signals with the seaweed method. For this we use a simulated base signal, seen in Figure 8, designed to emulate the signals observed in our bean beetle experiments, specifically the region of Figure 1 around frame 300. The equation for the base signal is given by

$$f(n) = \begin{cases} 0 & \text{if } 1 \leq n < \beta - 25, \\ 0.02n + 0.5 - 0.02\beta & \text{if } \beta - 25 \leq n < \beta - 10, \\ n + 10.3 - \beta & \text{if } \beta - 10 \leq n < \beta - 6, \\ 4.3 & \text{if } \beta - 6 \leq n < \beta, \\ -0.5n + 4.3 + 0.5\beta & \text{if } \beta \leq n < \beta + 10, \\ -0.25n + 1.8 + 0.25\beta & \text{if } \beta + 10 \leq n < \beta + 15, \\ -1.95 & \text{if } \beta + 15 \leq n < 70, \end{cases} \qquad (3\text{-}3)$$

where $\beta + 1$ is the initial large-scale drop in amplitude, which we call the primary transition point, and n is the time index. In Example 3.2 we use the seaweed method to identify the primary transition point in the base signal.
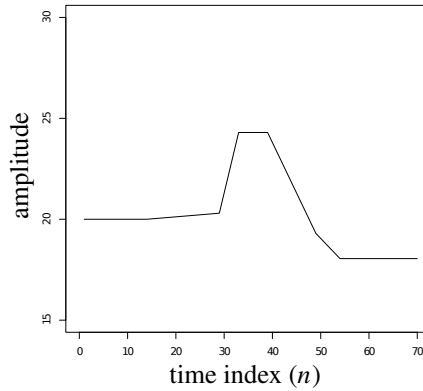
Figure 8. Base signal from (3-3) with primary transition point at $\beta + 1 = 40$.

Example 3.2. Consider the signal displayed in Figure 9. The primary transition point is located by following the long strand of solid-circle seaweed down. In following the right branch at the split, we end at the point $(n, 2j) = (40, 2)$, signifying that the transition occurs at $n = 40$. It is also interesting to note the stacks of triangle seaweed concentrated near the beginning and end of the signal. In each case, this behavior is the result of the constant amplitude of the signal, and the shape of these stacks is due to the changing size of the sliding window. Furthermore, in each case the stack is bordered by a strand of open-circle seaweed, which corresponds to the
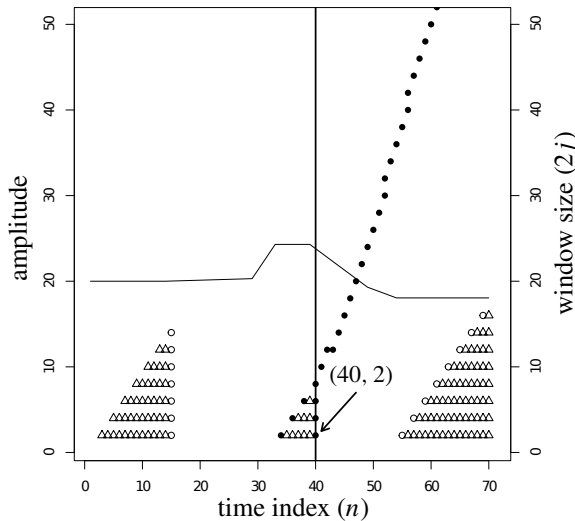
Figure 9. Base signal from (3-3) with primary transition point $\beta + 1 = 40$.
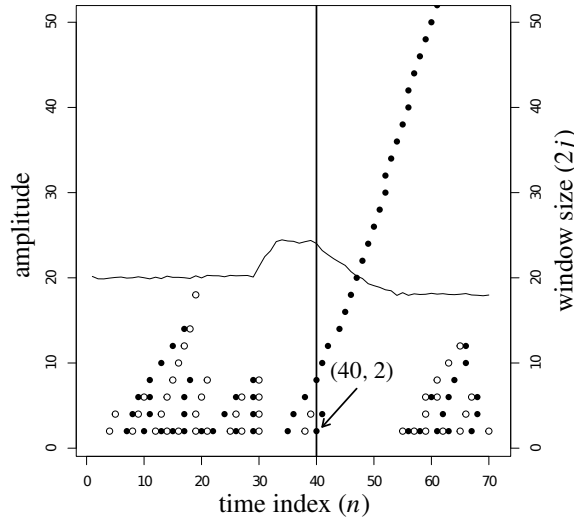
Figure 10. Example of a signal with low noise and limited seaweed branching. The primary transition point for this signal is $\beta + 1 = 40$, and the noise was sampled from a normal distribution with mean 0 and standard deviation 0.1.

first initial transition in the signal from constant to increasing, at $n = 15$, and the last transition in the signal from decreasing to constant, at $n = 55$.

We follow this example with another involving a slightly noisier signal with low levels of seaweed branching.

Example 3.3. Consider the signal displayed in Figure 10, which was obtained by adding noise to the base signal which was randomly sampled from a normal distribution with mean 0 and standard deviation 0.1. This example proceeds nearly identically to Example 3.2. We follow the largest strand of solid-circle seaweed down, following the right branch to the point $(n, 2j) = (40, 2)$. As before, this means that the primary transition point occurs at $n = 40$. Additionally, the window level corresponding to this point, $2j = 2$, indicates that the algorithm was able to recognize the transition point even at the highest possible resolution. Note that in this example, the presence of noise eliminates the constant behavior seen in Figure 9, and therefore we do not observe any open triangles being plotted in the output. Instead, we see branching strands of open- and closed-circle seaweed which appear on either side of localized perturbations due to noise.

Next, we consider an example with more noise and seaweed branching.

Example 3.4. Consider the signal displayed in Figure 11, which was obtained by adding noise to the base signal which was randomly sampled from a normal distribution with mean 0 and standard deviation 1.2. We begin again as in Example 3.2;
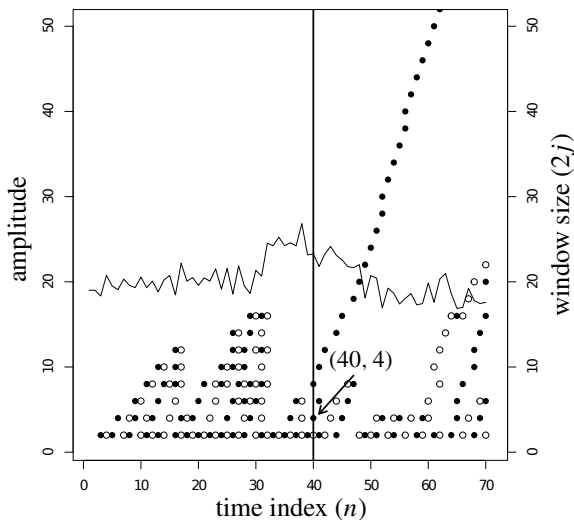
Figure 11. Example of a signal with a higher level of noise and seaweed branching. The primary transition point for this signal is $\beta + 1 = 40$, and the noise was sampled from a normal distribution with mean 0 and standard deviation 1.2.

however, in this case we can only follow the right strand to $(n, 2j) = (40, 4)$, as the seaweed below this point is focusing on high-frequency noise. Therefore, the primary transition occurs at $n = 40$.

3E. *Evaluating the seaweed method.* We designed experiments to test the accuracy and precision of the seaweed method, using the base signal shown in Figure 8 and defined in (3-3). The first experiment explores a broad range of noise levels with a small number of signals at each level. The second experiment explores a smaller range of noise levels with a larger number of signals at each level.

3E.1. *Experiment 1.* Noise was generated by randomly sampling from normal distributions with mean 0 and standard deviations 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2. We injected this noise into base signals by adding it to base signal. The primary transition point, $\beta + 1$, had been randomly selected from the interval $26 \le n \le 56$. Examples of this sort of signal can be seen in Examples 3.2, 3.3, and 3.4.

For each of the eleven standard deviations, 30 signals were generated with randomly generated noise and a randomly positioned primary transition point. Each of the 330 signals was stored in a file in random order, and the true primary transition point and noise level for each were stored in a separate file. This second file was kept hidden while we used seaweed plots to identify the primary transition points in the synthetic signals. These primary transition points were then compared to those in the

hidden file. The script seaweed_test1.R used to generate the signals and guarantee the conditions required for a blind trial is included in the online supplement.

3E.2. *Experiment 2.* Noise with standard deviations of 0.1, 0.2, 0.3, 0.4, 0.5 was injected into the base signal. Additionally, in this experiment we generated 200 signals for each standard deviation. With the exception of these two changes, everything else proceeded as in the first experiment. The script seaweed_test2.R for this experiment is also included in the online supplement.

4. Results

4A. *Experiment 1.* The difference between the true primary transition points and the primary transition points identified by the seaweed method were computed for each signal, and the error distributions that resulted for each level of noise were examined. The standard deviations of errors for Experiment 1 are displayed in Table 3.

As can be seen from Table 3, the noisier the signal the greater the error in identifying primary transition points. Furthermore, as can be seen in Figure 12, this trend is approximately linear, with $r^2 = 0.728$.

4B. *Experiment 2.* As in Experiment 1, the primary transition point identification error and standard deviations for the resulting error distributions were calculated for Experiment 2. These standard deviations are displayed in Table 4.

Again, it is apparent that as we add more noise to the signal, we observe more error in the analysis. As can be seen in Figure 12, this trend is also approximately linear, with $r^2 = 0.825$. Histograms of each error distribution in this experiment are displayed in Figure 13. Each has a mode at 0 and exhibits a higher standard deviation of primary transition point identification error as the noise level increases.

4C. *A bean beetle example.* Recall the time signal for the relative brightness of a bean beetle egg shown in Figure 1. This time signal is based on time-lapse

noise s.d.	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
error s.d. (time steps)	0	1.106	1.375	1.137	1.574	1.907	2.369	2.128	1.846	2.132	2.273

Table 3. Standard deviation of primary transition point identification error for each of the 11 levels of noise tested in Experiment 1.

noise s.d.	0.1	0.2	0.3	0.4	0.5
error s.d. (time steps)	0.615	0.808	0.845	0.990	0.941

Table 4. Standard deviation of primary transition point identification error for each of the 5 levels of noise tested in Experiment 2.
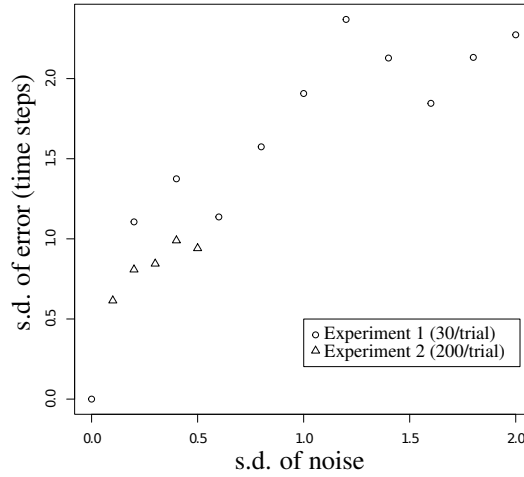
Figure 12. Scatterplot showing standard deviation of primary transition point identification error vs. standard deviation of noise for both experiments, showing their approximately linear relationship with $r^2 = 0.728$ for Experiment 1 and $r^2 = 0.825$ for Experiment 2.
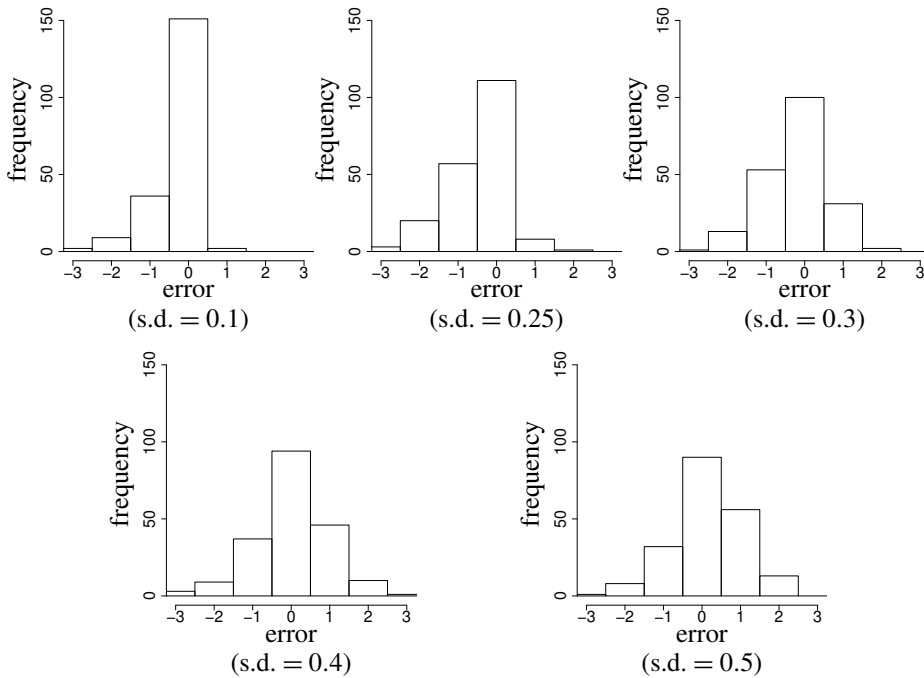
Figure 13. Histograms of error distributions in Experiment 2. All have a mode at 0 and exhibit a higher standard deviation of primary transition point identification error as the noise level increases.
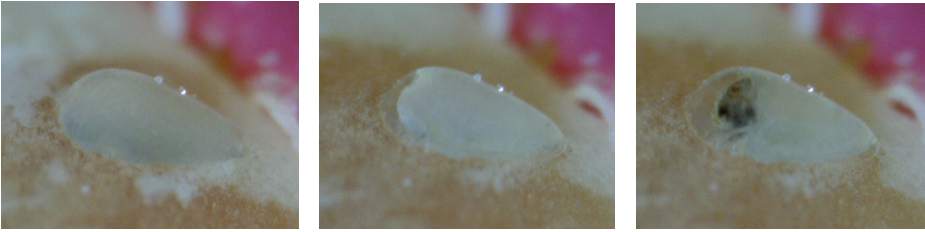
Figure 14. Digital microscope images of an egg of a developing bean beetle embryo. The image on the left was taken shortly after the egg was laid, the image in the middle was taken near the end of embryonic development, and the third image shows the embryo's darkened head capsule.

photographs of bean beetle eggs such as the ones shown in Figure 14. We used the seaweed method to analyze this signal, and determine the frame number associated with the beginning of head-capsule darkening. The signal with seaweed points is shown in Figure 15. In this case we used the seaweed method to select the point $(n, 2j) = (318, 4)$, meaning that the embryo's head capsule began to darken in frame 318. Furthermore, we know from experimental observations that this egg was laid in frame 93, which combined with the fact that the frames are 20 minutes apart, means that the embryo developed in approximately 3.125 days.
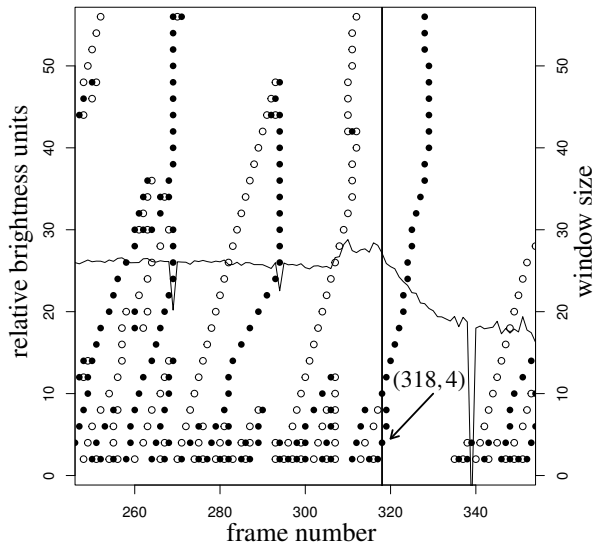
Figure 15. Graph of relative brightness of a particular bean beetle egg over time between frames 250 and 350 with seaweed overlaid. All of the frames are 20 minutes apart.
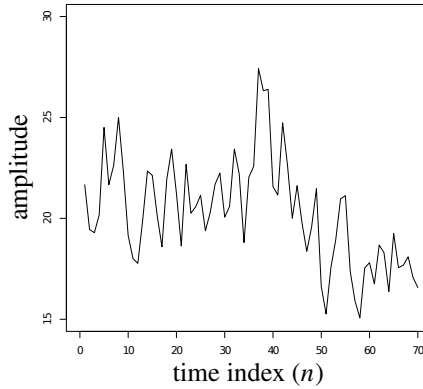
Figure 16. Base signal injected with noise sampled from a normal distribution with mean 0 and standard deviation 2, with primary transition point at $\beta + 1 = 40$.

5. Discussion

5A. *Experiment 1.* Despite the lower resolution with which we examined varying noise levels in Experiment 1, we were still able to glean a general idea of how the seaweed method performs at a wide range of noise. The correlation observed between noise level and primary transition point identification error seen in Table 3 and Figure 12 suggests that the seaweed method is most accurate at low levels of noise. However, even at higher levels of noise, such as noise with a standard deviation of 2, when it is difficult to visually recognize the presence of a signal beneath the noise, our method was still able to detect the primary transition point. An example of such a signal can be seen in Figure 16. For this level of noise, the resulting error distribution had a standard deviation of 2.273, meaning that most of the time the seaweed method is able to accurately identify the primary transition point within three time steps, even at such an extreme level of noise.

For lower levels of noise the standard deviations of the error distributions were closer to 1.5, meaning that at lower levels of noise this experiment suggested that the method is generally capable of accurately identifying the primary transition point within two time steps, or 40 minutes in our bean beetle experiments.

5B. *Experiment 2.* In Experiment 2 we chose to focus more closely on the range of noise with standard deviations around 0.2, the level of noise observed in the time signals from our bean beetle experiments. As can be seen from Table 4 and Figure 12, the approximately linear trend observed in Experiment 1 is also apparent at this scale, lending support to the conclusion that as more noise is introduced to the signal, more error is apparent in detection of primary transition points. Additionally, the increased sample size in this experiment resulted in a better representation of the

standard deviations of primary transition point identification error for this method, suggesting that the standard deviation of this error for low levels of noise is closer to one time step, or 20 minutes in our bean beetle experiments.

These conclusions are further reinforced through the histograms in Figure 13. All histograms exhibit a mode at 0 and standard deviations between 0 and 0.941, meaning that at the levels of noise examined in this experiment, the seaweed method is particularly accurate at identifying primary transition points. As the level of noise increases, the standard deviation of primary transition point identification error increases slightly, although the majority of the values fall between ± 1, meaning that our method is rarely more than one time step away in identifying the primary transition point for those levels of noise. Furthermore, none of the histograms include values more extreme than ± 3, meaning that our method is very rarely more than three time steps away from identifying the primary transition point.

We also observed a slight skew in the error distributions, in the form of a right skew at lower levels of noise and a left skew at higher levels of noise. The right skew may be the result of the asymmetry in the base signal, as seen in Figure 8, or potentially human subjectivity. At low levels of noise there may be a tendency to select the transition point without fully relying on the seaweed. This effect ought to be negligible, however, due to the manner in which the experiment was conducted. The left skew may be due to the fact that the algorithm selects the right endpoint of the window being considered and therefore may occasionally overshoot the transition point. Further experimentation, such as a replication of these experiments with a different base signal, would be needed to identify the source of the skew.

5C. *Conclusions.* The error distributions found in Experiments 1 and 2 have a mode at 0 and low standard deviations of primary transition point identification error, from which we can conclude that the seaweed method is both accurate and precise in identifying transition points. Although we have focused our discussion on transitions from positive to negative averages, our results should generalize to detecting other types of transition points. We have used the seaweed method to detect transitions from negative to positive averages to identify other biological markers in bean beetle development as well.

The conceptual transparency inherent to this method gives it the advantage of being approachable by a wide audience, including those without any substantial background in wavelets or signal processing. Its unique graphical representation of transition points also makes it easy to learn and apply. Although this may seem to imply a level of subjectivity, this can be avoided though careful adherence to the guidelines laid out in Section 3D.

This method seeks to identify transition points whose influence reaches across multiple time scales, localizing transitions of global importance. In addition, the

seaweed method provides a higher resolution of detecting these types of transitions when compared to the discrete wavelet transform.

The efficacy of this method also serves to support our analysis of bean beetle development times, as this method was essential in determining these times from signals like those discussed in Example 3.1. Furthermore, we believe that this method will prove capable of analyzing a much wider range of time signals, such as stock market data and audio signals, and may be used to analyze other types of signals (e.g., spatial) as well. One such spatial signal that could benefit from this type of analysis is population density of a given organism along a transect.

Acknowledgments

References

[Aboufadel and Schlicker 1999] E. Aboufadel and S. Schlicker, *Discovering wavelets*, Wiley, New York, NY, 1999. MR Zbl

[Auscher 1992] P. Auscher, "Wavelet bases for $L^2(\mathbb{R})$ with rational dilation factor", pp. 439–451 in *Wavelets and their applications*, edited by M. B. Ruskai et al., Jones and Bartlett, Boston, MA, 1992. MR Zbl

[Bayram and Selesnick 2009] İ. Bayram and I. W. Selesnick, "Overcomplete discrete wavelet transforms with rational dilation factors", *IEEE Trans. Signal Process.* **57**:1 (2009), 131–145. MR

[Beck and Blumer 2011] C. W. Beck and L. S. Blumer, "A handbook on bean beetles, *Callosobruchus maculatus*", 2011, available at http://www.beanbeetles.org/handbook/handbook.pdf.

[Bénéteau and Van Fleet 2011] C. Bénéteau and P. J. Van Fleet, "Discrete wavelet transformations and undergraduate education", *Notices Amer. Math. Soc.* **58**:5 (2011), 656–666. MR

[Christensen 2001] O. Christensen, "Frames, Riesz bases, and discrete Gabor/wavelet expansions", *Bull. Amer. Math. Soc. (N.S.)* **38**:3 (2001), 273–291. MR Zbl

[Daubechies et al. 2003] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: MRA-based constructions of wavelet frames", *Appl. Comput. Harmon. Anal.* **14**:1 (2003), 1–46. MR Zbl

[Logan and Powell 2001] J. A. Logan and J. A. Powell, "Ghost forests, global warming and the mountain pine beetle", *Amer. Entomol.* **47**:3 (2001), 160–173.

[Logan et al. 1976] J. A. Logan, D. J. Wollkind, S. C. Hoyt, and L. K. Tanigoshi, "An analytic model for description of temperature-dependent phenomena in arthropods", *Environ. Entomol.* **5**:6 (1976), 1133–1140.

[Mulcahy 1996] C. Mulcahy, "Plotting and scheming with wavelets", *Math. Mag.* **69**:5 (1996), 323–343. MR Zbl

[Selesnick 2011] I. W. Selesnick, "Wavelet transform with tunable Q-factor", *IEEE Trans. Signal Process.* **59**:8 (2011), 3560–3575. MR

[Teolis 1998] A. Teolis, *Computational signal processing with wavelets*, Birkhäuser, Boston, MA, 1998. MR Zbl

[Yurk and Powell 2009] B. Yurk and J. A. Powell, "Modeling the evolution of insect phenology", *Bull. Math. Biol.* **71**:4 (2009), 952–979. MR Zbl

david.mcmorris@huskers.unl.edu

                         Department of Mathematics, University of Nebraska–Lincoln,
                         203 Avery Hall, PO BOX 880130, Lincoln, NE 68588,
                         United States

pearsonp@hope.edu   *Department of Mathematics, Hope College, 27 Graves Place,*
                         Holland, MI 49422, United States

yurk@hope.edu   *Department of Mathematics, Hope College, 27 Graves Place,*
                         Holland, MI 49423, United States

msp

A generalization of the matrix transpose map and its relationship to the twist of the polynomial ring by an automorphism

Andrew McGinnis and Michaela Vancliff

(Communicated by Vadim Ponomarenko)

A generalization of the notion of symmetric matrix was introduced by Cassidy and Vancliff in 2010 and used by them in a construction that produces quadratic regular algebras of finite global dimension that are generalizations of graded Clifford algebras. In this article, we further their ideas by introducing a generalization of the matrix transpose map and use it to generalize the notion of skew-symmetric matrix. With these definitions, an analogue of the result that every $n \times n$ matrix is a sum of a symmetric matrix and a skew-symmetric matrix holds. We also prove an analogue of the result that the transpose map is an antiautomorphism of the algebra of $n \times n$ matrices, and show that the antiautomorphism property of our generalized transpose map is related to the notion of twisting the polynomial ring on n variables by an automorphism.

Introduction

In [Cassidy and Vancliff 2010], a generalization of the notion of symmetric matrix was introduced and used in a construction that produces quadratic regular algebras of finite global dimension that are generalizations of graded Clifford algebras. In the same paper, it was also shown that such a matrix corresponds to a noncommutative analogue of a quadratic form. In this article, we further these ideas by introducing a generalization of the matrix transpose map and use it to generalize the notion of skew-symmetric matrix. In particular, we prove in Theorem 2.5 an analogue of the result that every $n \times n$ matrix is a sum of a symmetric matrix and a skew-symmetric matrix. We also prove, in Proposition 2.6 and Corollary 2.16, an analogue of the result that the transpose map is an antiautomorphism of the algebra of $n \times n$ matrices. This latter property is shown in Corollary 2.16 to be related to the twist of the polynomial ring on n variables by an automorphism.

The article is outlined as follows. In Section 1, we define generalizations of symmetric and skew-symmetric matrices together with a few other concepts that will be used in the subsequent section. Section 2 is in two parts: the first defines and explores a generalization of the transpose map, whereas the second ties the behavior of this transpose map to the notion of twisting a polynomial ring by an automorphism.

1. Definitions

In this section, we recall the generalizations of symmetric matrix and quadratic form that were introduced in [Cassidy and Vancliff 2010]. We also introduce a generalization of the notion of skew-symmetric matrix.

Throughout, $\Bbbk$ denotes a field. We use the notation $M(n, \Bbbk)$ to denote the vector space of $n \times n$ matrices with entries in $\Bbbk$ and $M(m, r, \Bbbk)$ to denote the vector space of $m \times r$ matrices with entries in $\Bbbk$. For any matrix $N \in M(m, r, \Bbbk)$, we let N_{ij} denote the ij-entry of N.

Definition 1.1. Let $\mu \in M(n, \Bbbk)$ be such that $\mu_{ij}\mu_{ji} = 1$ for all distinct i, j. A matrix $M \in M(n, \Bbbk)$ is said to be

(a) μ-symmetric if $M_{ij} = \mu_{ij}M_{ji}$ for all i, j [Cassidy and Vancliff 2010];

(b) skew-μ-symmetric if $M_{ij} = -\mu_{ij}M_{ji}$ for all i, j.

If $\mu_{ij} = 1$ for all i, j, then any μ-symmetric matrix is a symmetric matrix, and any skew-μ-symmetric matrix is a skew-symmetric matrix. Consequently, we generalize the notion of transpose in the next section and relate the notions of μ-symmetry and skew-μ-symmetry to that concept.

The notion of μ-symmetry was used in [Cassidy and Vancliff 2010] to produce algebras that may be viewed as quantized graded Clifford algebras. In other words, the main use of μ-symmetry is to "tie together" two or more matrices to a particular matrix μ, and to do so in a symmetrical manner.

Following [Vancliff and Veerapen 2013], we write $M^\mu(n, \Bbbk)$ for the set of μ-symmetric $n \times n$ matrices with entries in $\Bbbk$. Likewise, we write $M^{s\mu}(n, \Bbbk)$ for the set of skew-μ-symmetric $n \times n$ matrices with entries in $\Bbbk$. Clearly, $M^\mu(n, \Bbbk)$ and $M^{s\mu}(n, \Bbbk)$ are subspaces of $M(n, \Bbbk)$.

Mirroring the theory for symmetric matrices and following [Cassidy and Vancliff 2010], a μ-symmetric matrix corresponds to a noncommutative analogue of a quadratic form, provided $\mu_{ii} = 1$ for all i; this correspondence is summarized as follows.

Definition 1.2 [Cassidy and Vancliff 2010]. Let $\mu \in M(n, \Bbbk)$ be as in Definition 1.1, with the additional assumption that $\mu_{ii} = 1$ for all i. Let (S, μ) denote the quadratic $\Bbbk$-algebra on generators $z_1, \ldots, z_n$ with defining relations $z_jz_i = \mu_{ij}z_iz_j$ for all $i, j = 1, \ldots, n$, and let S_2 denote the span of the homogeneous elements of (S, μ) of degree two. A (noncommutative) quadratic form is defined to be any element of S_2.

The algebra (S, μ) has no zero divisors and has the same Hilbert series as the polynomial ring on n variables. By [Cassidy and Vancliff 2010], if $\mu_{ii} = 1$ for all i, then $M^\mu(n, \Bbbk) \cong S_2$, as vector spaces, via the map $M \mapsto z^T M z \in S_2$, where $z = (z_1, \ldots, z_n)^T$.

In the next section, the algebra (S, μ) will be considered in the special case where $\mu_{ij} = \mu_{ik}\mu_{kj}$ for all $i, j, k = 1, \ldots, n$. By [Nafari and Vancliff 2015, Lemma 2.2], (S, μ) is a twist (see Definition 1.3 below) of the polynomial ring R on n variables by a graded automorphism of R of degree zero if and only if this condition on μ holds.

Definition 1.3 [Artin et al. 1991, §8]. Let $A = \bigoplus_{k \geq 0} A_k$ be a graded $\Bbbk$-algebra and let ϕ be a graded degree-zero automorphism of A. The twist A' of A by ϕ is a graded $\Bbbk$-algebra that is the vector space $\bigoplus_{k \geq 0} A_k$ with a new multiplication $*$ defined as follows: if $a' \in A'_i = A_i$ and $b' \in A'_j = A_j$, then $a' * b' = (a\phi^i(b))'$, where the right-hand side is computed using the original multiplication in A and a, b are the images of a', b', respectively, in A.

Clearly, the twist of a quadratic algebra is again a quadratic algebra. Moreover, this notion of twist is reflexive and symmetric.

2. Main results

In this section, we define a generalization of the notion of transpose of a matrix and explore properties of this new concept. Our main results are given in Theorem 2.5, Proposition 2.6, Theorem 2.15 and Corollary 2.16.

2A. *The transpose map.*

Definition 2.1. If $\nu \in M(r, m, \Bbbk)$ and $N \in M(m, r, \Bbbk)$, we define the ν-transpose of N, denoted $N^{\nu T}$, to be the $r \times m$ matrix with ij-entry given by $\nu_{ij}N_{ji}$ for all i, j.

Clearly, if $\nu_{ij} = 1$ for all i, j, then the ν-transpose map is the transpose map. Alternatively, we may view the ν-transpose as a composition of maps; for this purpose, let $\hat{\nu} : M(r, m, \Bbbk) \to M(r, m, \Bbbk)$ be defined by $\hat{\nu}(K) = (\nu_{ij}k_{ij})$, where $K = (k_{ij}) \in M(r, m, \Bbbk)$.

Lemma 2.2. *If ν, $\hat{\nu}$ and N are as above, then $N^{\nu T} = \hat{\nu}(N^T)$, where N^T denotes the transpose of N. In particular, the ν-transpose map is a linear transformation.* $\square$

Lemma 2.3. *Let μ be as in Definition 1.1. A matrix $M \in M(n, \Bbbk)$ is μ-symmetric if and only if $M^{\mu T} = M$. Additionally, M is skew-μ-symmetric if and only if $M^{\mu T} = -M$.*

Proof. If $M \in M(n, \Bbbk)$ is μ-symmetric, then $M_{ij} = \mu_{ij}M_{ji}$ for all i, j, so $M = M^{\mu T}$; reversing the argument proves the converse. The proof of skew-μ-symmetric case is similar. $\square$

Proposition 2.4. *Let $\mu \in M(n, \Bbbk)$ be such that $\mu_{ij}\mu_{ji} = 1$ for all i, j. If $M \in M(n, \Bbbk)$, then*

(a) $(M^{\mu T})^{\mu T} = M$,

(b) $M + M^{\mu T} \in M^{\mu}(n, \Bbbk)$,

(c) $M - M^{\mu T} \in M^{s\mu}(n, \Bbbk)$.

Proof. (a) We have $[M^{\mu T}]^{\mu T} = (\mu_{ij}M_{ji})^{\mu T} = (\mu_{ij}\mu_{ji}M_{ij}) = (M_{ij}) = M$.

(b)–(c) We have $M \pm M^{\mu T} = (M_{ij} \pm \mu_{ij}M_{ji}) = \left(\pm\mu_{ij}(M_{ji} \pm \mu_{ji}M_{ij})\right)$. Thus,

$$[M \pm M^{\mu T}]^{\mu T} = \left(\pm\mu_{ij}\mu_{ji}(M_{ij} \pm \mu_{ij}M_{ji})\right) = \left(\pm(M_{ij} \pm \mu_{ij}M_{ji})\right) = \pm[M \pm M^{\mu T}],$$

and so the result follows from Lemma 2.3. $\square$

Theorem 2.5. *Suppose $\mathrm{char}(\Bbbk) \neq 2$. If $\mu \in M(n, \Bbbk)$ is such that $\mu_{ij}\mu_{ji} = 1$ for all i, j, then*

$$M(n, \Bbbk) = M^{\mu}(n, \Bbbk) \oplus M^{s\mu}(n, \Bbbk).$$

Proof. If $M \in M(n, \Bbbk)$, then $M = \frac{1}{2}(M + M^{\mu T}) + \frac{1}{2}(M - M^{\mu T})$, since $\mathrm{char}(\Bbbk) \neq 2$. It follows from Proposition 2.4 that $M(n, \Bbbk) = M^{\mu}(n, \Bbbk) + M^{s\mu}(n, \Bbbk)$. However, the assumption on the characteristic of $\Bbbk$ ensures that $M^{\mu}(n, \Bbbk) \cap M^{s\mu}(n, \Bbbk) = \{0\}$, which completes the proof. $\square$

A well-known result for symmetric matrices is that if $X \in M(n, \Bbbk)$ is symmetric, then $P^T X P$ is also symmetric for all $P \in M(n, \Bbbk)$. This result is a consequence of the fact that $[XY]^T = Y^T X^T$ for all $X, Y \in M(n, \Bbbk)$; that is, the transpose map is an antiautomorphism of $M(n, \Bbbk)$. However, the analogues of these results are false in general for μ-symmetry, unless μ satisfies certain conditions as follows.

Proposition 2.6. *If $\mu \in M(n, \Bbbk)$ is such that $\mu_{ij} = \mu_{ik}\mu_{kj}$ for all i, j, k, then $[XY]^{\mu T} = Y^{\mu T} X^{\mu T}$ for all $X, Y \in M(n, \Bbbk)$.*

Proof. Let $X, Y \in M(n, \Bbbk)$. We have

$$[XY]^{\mu T} = \left(\sum_{k=1}^{n} X_{ik}Y_{kj}\right)^{\mu T} = \left(\mu_{ij}\sum_{k=1}^{n} X_{jk}Y_{ki}\right),$$

whereas

$$Y^{\mu T} X^{\mu T} = (\mu_{ik}Y_{ki})(\mu_{kj}X_{jk}) = \left(\sum_{k=1}^{n} \mu_{ik}\mu_{kj}Y_{ki}X_{jk}\right) = \left(\mu_{ij}\sum_{k=1}^{n} X_{jk}Y_{ki}\right),$$

where the last equality is a consequence of the condition on μ. $\square$

If $\mu \in M(n, \Bbbk)$ satisfies the hypotheses of Propositions 2.4 and 2.6, then $\mu_{ij} = \mu_{ik}\mu_{kj}$ for all i, j, k, and $\mu_{ii} = 1$ for all i; the converse also holds.

Corollary 2.7. *Let $\mu \in M(n, \Bbbk)$. If $\mu_{ij} = \mu_{ik}\mu_{kj}$ for all i, j, k, and if $\mu_{ii} = 1$ for all i, then $P^{\mu T} X P \in M^\mu(n, \Bbbk)$ for all $X \in M^\mu(n, \Bbbk)$ and for all $P \in M(n, \Bbbk)$.*

Proof. The conditions on μ imply $\mu_{ik}\mu_{ki} = \mu_{ii} = 1$ for all i, k, so that Lemma 2.3 and Propositions 2.4 and 2.6 may be applied to compute $[P^{\mu T} X P]^{\mu T}$; namely,

$$[P^{\mu T} X P]^{\mu T} = P^{\mu T}[P^{\mu T} X]^{\mu T} = P^{\mu T} X^{\mu T} [P^{\mu T}]^{\mu T} = P^{\mu T} X P$$

for all $X \in M^\mu(n, \Bbbk)$ and for all $P \in M(n, \Bbbk)$. The result follows from Lemma 2.3. $\square$

The hypotheses on μ in the last result coincide with the hypotheses required for the skew polynomial ring (S, μ), defined in Definition 1.2, to be a twist (in the sense of Definition 1.3) of the polynomial ring R on n variables by a graded automorphism of R of degree zero. However, the above methods give no insight as to why this should be the case, so further analysis is required to explain this relationship and is the purpose of the next subsection.

2B. *The transpose map and twisting the polynomial ring.* The goal of this subsection is to show that the result of Corollary 2.7 is directly related to the algebra (S, μ) being a twist of the polynomial ring R as mentioned at the end of Section 2A. Our method will be to show that the result of Corollary 2.7 is directly related to a certain map $\bar{\mu} : M(n, \Bbbk) \to M(n, \Bbbk)$ (see Definition 2.14) being an automorphism, in which case $\bar{\mu}$ induces an automorphism of (S, μ) that twists (S, μ) to R.

Throughout this subsection, we assume that $\mu_{ii} = 1$ for all i and that $\mu_{ij}\mu_{ji} = 1$ for all i, j.

Let V denote the span of the homogeneous elements of (S, μ) of degree one. Since (S, μ) is a domain, for each $k = 1, \ldots, n$, we may define $\theta_k \in \mathrm{Aut}(S, \mu)$ via $sz_k = z_k\theta_k(s)$ for all $s \in (S, \mu)$. In particular, for every k, we have $\theta_k(z_i) = \mu_{ki}z_i$ for all i, so if we twist (S, μ) by θ_k, we obtain a quadratic algebra in which the image of z_k is central.

Let V^* denote the vector-space dual of V and let $\{z_1^*, \ldots, z_n^*\}$ in V^* denote the dual basis to the basis $\{z_1, \ldots, z_n\}$ of V. For each k, the linear transformation $\theta_k|_V : V \to V$ induces a linear map $\theta_k^* : V^* \to V^*$, where $\theta_k^*(z_i^*) = \mu_{ik}z_i^*$ for all i. Hence θ_k induces a linear map $\bar{\theta}_k : V \otimes_{\Bbbk} V^* \to V \otimes_{\Bbbk} V^*$ via

$$\bar{\theta}_k(v \otimes u) = \theta_k(v) \otimes \theta_k^*(u)$$

for all $v \otimes u \in V \otimes_{\Bbbk} V^*$.

Remark 2.8. As is well known, $V \otimes_{\Bbbk} V^*$ is a $\Bbbk$-algebra under the usual addition and with multiplication given by $(v \otimes u)(v' \otimes u') = (uv')(v \otimes u')$ for all $v, v' \in V$, $u, u' \in V^*$. In fact, $V \otimes_{\Bbbk} V^* \cong M(n, \Bbbk)$, as $\Bbbk$-algebras, via the map that sends $z_i \otimes z_j^*$ to the $n \times n$ matrix with 1 in the ij-entry and zeros elsewhere.

Lemma 2.9. *For every $k = 1, \ldots, n$, the linear map $\bar{\theta}_k$ is in $\mathrm{Aut}(V \otimes_{\Bbbk} V^*)$.*

Proof. Since $\bar{\theta}_k$ is linear and bijective, it remains to prove that $\bar{\theta}_k$ respects multiplication, and it suffices to consider products of pure tensors. Let $v, v' \in V$ and $u, u' \in V^*$, and write $v' = \sum_{i=1}^n v_i z_i$ and $u = \sum_{j=1}^n u_j z_j^*$, where $v_i, u_j \in \Bbbk$ for all i, j. In particular, $uv' = \sum_{i=1}^n u_i v_i$ and

$$\theta_k^*(u)\theta_k(v') = \left(\sum_{j=1}^n u_j \mu_{jk} z_j^* \right)\left(\sum_{i=1}^n v_i \mu_{ki} z_i \right) = \sum_{i=1}^n u_i v_i = uv'.$$

It follows that
$$\bar{\theta}_k\big((v \otimes u)(v' \otimes u')\big) = \bar{\theta}_k\big((uv')(v \otimes u')\big)$$
$$= uv'\theta_k(v) \otimes \theta_k^*(u'),$$

whereas
$$\bar{\theta}_k(v \otimes u)\bar{\theta}_k(v' \otimes u') = \big(\theta_k(v) \otimes \theta_k^*(u)\big)\big(\theta_k(v') \otimes \theta_k^*(u')\big)$$
$$= \theta_k^*(u)\theta_k(v')\big(\theta_k(v) \otimes \theta_k^*(u')\big),$$

so the result follows. □

In the following, $\Bbbk^\times$ denotes the nonzero elements of $\Bbbk$.

Lemma 2.10. *For all k, i, we have $\bar{\theta}_k = \bar{\theta}_i$ if and only if $\theta_k \in \Bbbk^\times \theta_i$.*

Proof. We have $\theta_k = \lambda \theta_i$ for some $\lambda \in \Bbbk^\times$ if and only if $\theta_k^* = \lambda^{-1}\theta_i^*$. The result follows from the definitions of $\bar{\theta}_k$ and $\bar{\theta}_i$. □

Proposition 2.11. *The map θ_k is in $\Bbbk^\times \theta_1$ for all k if and only if the algebra (S, μ) is a twist (in the sense of Definition 1.3) of the polynomial ring on n variables.*

Proof. As mentioned above, for each k, the twist of (S, μ) by θ_k yields an algebra in which the image of z_k is central. Hence, if $\theta_k \in \Bbbk^\times \theta_1$ for all k, then twisting by θ_k produces an algebra R in which the image of z_i is central for all i. Since the relations of R are induced by the relations of (S, μ), it follows that R is the polynomial ring on n variables.

Conversely, suppose (S, μ) is a twist of the polynomial ring R on n variables. It follows that there exists a degree-zero map $\theta \in \mathrm{Aut}(S, \mu)$ such that twisting (S, μ) by θ renders the image of z_k central in R for all k. Writing "$\cdot$" for the multiplication in R, this implies

$$z_k \theta(z_i) = z_k \cdot z_i = z_i \cdot z_k = z_i \theta(z_k)$$

for all i, k. However, since S is a quadratic algebra and since S_2 has a $\Bbbk$-basis $\{z_j z_l : 1 \le j \le l \le n\}$, it follows that $\theta(z_k) \in \Bbbk^\times z_k$ for all k. Writing $\theta(z_k) = \lambda_k z_k$, where $\lambda_k \in \Bbbk^\times$ for all k, we have $\mu_{ik} = \lambda_k/\lambda_i$ for all i, k and $\lambda_i \theta_i = \theta$ for all i. Thus, $\theta_k \in \Bbbk^\times \theta_1$ for all k. □

Corollary 2.12. *We have $\bar{\theta}_k = \bar{\theta}_1$ for all k if and only if (S, μ) is a twist (in the sense of Definition 1.3) of the polynomial ring on n variables.*

Proof. The result follows by combining Lemma 2.10 with Proposition 2.11. □

Lemma 2.13. *If $\bar{\theta}_k = \bar{\theta}_1$ for all k, then $\bar{\theta}_k((a_{ij})) = (\mu_{ji}a_{ij})$ for all k and for all $(a_{ij}) \in M(n, \Bbbk)$, where $M(n, \Bbbk)$ is identified with $V \otimes_{\Bbbk} V^*$ as in Remark 2.8.*

Proof. By identifying $M(n, \Bbbk)$ with $V \otimes_{\Bbbk} V^*$, we may write $(a_{ij}) \in M(n, \Bbbk)$ as

$$(a_{ij}) = \left(z_1 \otimes \sum_{i=1}^{n} a_{1i} z_i^* \right) + \left(z_2 \otimes \sum_{i=1}^{n} a_{2i} z_i^* \right) + \cdots + \left(z_n \otimes \sum_{i=1}^{n} a_{ni} z_i^* \right).$$

If $\bar{\theta}_k = \bar{\theta}_1$ for all k, then

$$\bar{\theta}_k((a_{ij})) = \sum_{j=1}^{n} \bar{\theta}_k \left(z_j \otimes \sum_{i=1}^{n} a_{ji} z_i^* \right) = \sum_{j=1}^{n} \bar{\theta}_j \left(z_j \otimes \sum_{i=1}^{n} a_{ji} z_i^* \right)$$

$$= \sum_{j=1}^{n} \left(\theta_j(z_j) \otimes \sum_{i=1}^{n} a_{ji} \theta_j^*(z_i^*) \right) = \sum_{j=1}^{n} \left(z_j \otimes \sum_{i=1}^{n} \mu_{ij} a_{ji} z_i^* \right) = (\mu_{ji} a_{ij}). \quad \square$$

Lemma 2.13 motivates the following definition.

Definition 2.14. Define $\bar{\mu} : M(n, \Bbbk) \to M(n, \Bbbk)$ by $\bar{\mu}((a_{ij})) = (\mu_{ji} a_{ij})$ for all $(a_{ij}) \in M(n, \Bbbk)$.

Moreover, $\bar{\mu} = (\ )^T \circ \hat{\mu} \circ (\ )^T$, where $\hat{\mu}$ is defined just prior to Lemma 2.2. Clearly, $\bar{\mu}$ is linear; with the assumption on μ at the start of Section 2B, $\bar{\mu}$ is also invertible.

Theorem 2.15. *The map $\bar{\mu}$ is an automorphism of $M(n, \Bbbk)$ if and only if the algebra (S, μ) is a twist of the polynomial ring on n variables.*

Proof. Identify $M(n, \Bbbk)$ with $V \otimes V^*$ as in Remark 2.8, so that we may view $\bar{\mu} : V \otimes V^* \to V \otimes V^*$. In particular, $\bar{\mu}(z_i \otimes z_j^*) = \mu_{ji}(z_i \otimes z_j^*)$ for all i, j. If (S, μ) is a twist of the polynomial ring, then $\bar{\mu} = \bar{\theta}_k$ for all k by Corollary 2.12 and Lemma 2.13. Hence $\bar{\mu}$ is an automorphism by Lemma 2.9.

Conversely, suppose $\bar{\mu}$ is an automorphism. It follows that

$$\bar{\mu}\big((z_j \otimes z_k^*)(z_k \otimes z_i^*)\big) = \bar{\mu}(z_j \otimes z_k^*)\bar{\mu}(z_k \otimes z_i^*)$$

for all i, j, k. Hence,

$$\bar{\mu}\big(z_k^* z_k (z_j \otimes z_i^*)\big) = \mu_{kj}(z_j \otimes z_k^*)\mu_{ik}(z_k \otimes z_i^*)$$

for all i, j, k, so that we have

$$\mu_{ij}(z_j \otimes z_i^*) = \mu_{ik}\mu_{kj}(z_j \otimes z_i^*)$$

for all i, j, k. It follows that $\mu_{ij} = \mu_{ik}\mu_{kj}$ for all i, j, k, so that (S, μ) is a twist of the polynomial ring by [Nafari and Vancliff 2015, Lemma 2.2]. $\quad \square$

Corollary 2.16. *The algebra (S, μ) is a twist of the polynomial ring if and only if $[XY]^{\mu T} = Y^{\mu T} X^{\mu T}$ for all $X, Y \in M(n, \Bbbk)$.*

Proof. Identify $M(n, \Bbbk)$ with $V \otimes V^*$ as in Remark 2.8. Considering Definitions 2.1 and 2.14, $X^{\mu T} = [\bar{\mu}(X)]^T$ for all $X \in M(n, \Bbbk)$. By Theorem 2.15, (S, μ) is a twist of the polynomial ring if and only if $\bar{\mu}$ is an automorphism, that is, if and only if $\bar{\mu}(XY) = \bar{\mu}(X)\bar{\mu}(Y)$ for all $X, Y \in M(n, \Bbbk)$. However, this holds if and only if $[\bar{\mu}(XY)]^T = [\bar{\mu}(X)\bar{\mu}(Y)]^T = [\bar{\mu}(Y)]^T[\bar{\mu}(X)]^T$ for all $X, Y \in M(n, \Bbbk)$, that is, if and only if $[XY]^{\mu T} = Y^{\mu T} X^{\mu T}$ for all $X, Y \in M(n, \Bbbk)$. $\square$

In view of this last result, it is clearer why the technical condition on μ is required in Corollary 2.7; the insight is that $\bar{\mu}$ needs to be an automorphism in order to have the μ-transpose map be an antiautomorphism, but that condition on $\bar{\mu}$ allows n automorphisms of (S, μ) to "merge" into one automorphism (denoted θ in the proof of Proposition 2.11) that twists (S, μ) to the polynomial ring.

References

[Artin et al. 1991] M. Artin, J. Tate, and M. Van den Bergh, "Modules over regular algebras of dimension 3", *Invent. Math.* **106**:2 (1991), 335–388. MR Zbl

[Cassidy and Vancliff 2010] T. Cassidy and M. Vancliff, "Generalizations of graded Clifford algebras and of complete intersections", *J. Lond. Math. Soc.* (2) **81**:1 (2010), 91–112. MR Zbl

[Nafari and Vancliff 2015] M. Nafari and M. Vancliff, "Graded skew Clifford algebras that are twists of graded Clifford algebras", *Comm. Algebra* **43**:2 (2015), 719–725. MR Zbl

[Vancliff and Veerapen 2013] M. Vancliff and P. P. Veerapen, "Generalizing the notion of rank to noncommutative quadratic forms", pp. 241–250 in *Noncommutative birational geometry, representations and combinatorics*, edited by A. Berenstein and V. Retakh, Contemp. Math. **592**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

mcginnis82292@gmail.com   *Department of Mathematics, University of California at Riverside, Riverside, CA 92521, United States*

vancliff@uta.edu   *Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019, United States*

■msp

Mixing times for the rook's walk via path coupling

Cam McLeman, Peter T. Otto, John Rahmani and Matthew Sutter

(Communicated by John C. Wierman)

The mixing time of a convergent Markov chain measures the number of steps required for the state distribution to be within a prescribed distance of the stationary distribution. In this paper, we illustrate the strength of the probabilistic technique called coupling and its extension, path coupling, to bound the mixing time of Markov chains. The application studied is the rook's walk on an n^d-chessboard, for which the mixing time has recently been studied using the spectral method. Our path-coupling result improves the previously obtained spectral bounds and includes an asymptotically tight upper bound in n for the two-dimensional case.

1. Introduction

In the standard game of chess, a rook occupies one of the 8^2 squares of the board, and moves by translating any distance along the row or column of its current position.

Suppose one came across a chessboard as in Figure 1 and was told that two legal rook moves had just been completed. Clearly there is no way to definitively deduce from this scant information what the position of the rook had been before the two moves — any of the 64 squares are possibilities — but could one guess the correct square in such a way as to have better odds than $\frac{1}{64}$? Certainly! Of the $14^2 = 196$ two-move sequences that a rook can undertake from a given position, there are 14 that would return it to its original square, six more that would take it to any given square in the same row or column, and only two such sequences to take the rook to any of the 49 remaining squares. The optimal strategy, over four times better than guessing a square uniformly at random, is to guess that the two moves left the rook in the same square as it started.

In the language of probabilistic processes, we might say that the chessboard is *unmixed* after only two moves — the location of the rook after two moves is heavily influenced by its starting position. In many applications of such processes, one is interested in the time it takes — the *mixing time* — in order for the influence of the starting position to be marginalized. In our chessboard example, this would
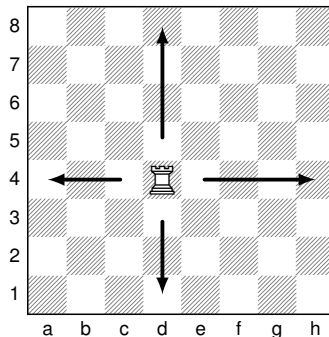
Figure 1. The 14 legal rook moves on a standard 8×8 chessboard.

be the number t such that after t random rook moves, its position is essentially governed by the uniform distribution. (The word "essentially" is needed here since the effect of the starting position never *completely* disappears — we insist only that the distribution be sufficiently close to the uniform distribution, using metrics introduced in the next section).

A standard approach to finding and bounding mixing times for processes like the rook's walk is via the spectral method: one writes down a 64×64 transition matrix encoding the squares of the chessboard and the probabilities of making moves between them, and one can then bound the mixing time in terms of the eigenvalues of this matrix. Such a method was used in [Kim 2012] to bound the mixing time of a rook's walk on the more general n^d-chessboard — a d-dimensional chessboard comprised of n squares in each dimension, on which a rook can move along precisely one dimension. Subsequently, in [Li and Tucker 2014], the exact mixing time for the rook's walk on n^2-chessboards was derived by direct computation, and the exact results showed how loose the spectral-method bounds were.

Here we take another approach to bounding mixing times from above. Namely, we use the *path coupling method* rather than spectral methods, leading to a much tighter upper bound. In Section 2, we introduce the basic theory of mixing times and Markov chains, including the metric one uses to measure the distance between distributions. This leads us to a formal definition of the mixing time. Section 3 specializes this material to the context of the rook's walk, and surveys the current state of knowledge concerning mixing times for the rook's walk. Sections 4 and 5 respectively introduce and implement path coupling in the context of the rook's walk, and Section 6 consists of an analysis and exhibition of the mixing times thus achieved.

2. Mixing times of Markov chains

The rook's walk studied in this paper is an example of a finite, discrete-time Markov chain, which, by definition, is a sequence of random variables $(X_t)_{t=0}^{\infty}$ that take

values in a finite state space Ω and satisfy the *Markov* (or memoryless) property

$$\mathbf{P}(X_{t+1}=y \mid X_0=x_0, X_1=x_1, \ldots, X_t=x) = \mathbf{P}(X_{t+1}=y \mid X_t=x).$$

The Markov property means that what happens next (i.e., in the $(t+1)$-th step) only depends on the state of the current step ($X_t = x$), and not the previous steps. Consequently, Markov chains are fully defined by their one-step transition probabilities $\mathbf{P}(X_{t+1}=y \mid X_t=x)$. We will also only be interested in time-homogeneous Markov chains, i.e., those for which

$$\mathbf{P}(X_{t+1}=y \mid X_t=x) = \mathbf{P}(X_{s+1}=y \mid X_s = x) \quad \text{for all } s, t.$$

Therefore, when discussing the one-step transition probabilities, one can just consider the case $t = 0$, that is, $\mathbf{P}(X_1=y \mid X_0=x)$. See [Kemeny et al. 1976] for a reference on the general theory of Markov chains.

The $|\Omega| \times |\Omega|$ matrix P consisting of entries $P(x, y) = \mathbf{P}(X_1 = y \mid X_0=x)$ is called the (one-step) *transition matrix* of the Markov chain. Note that P is a stochastic matrix which means that its entries are nonnegative and

$$\sum_{y \in \Omega} P(x, y) = 1 \quad \text{for all } x \in \Omega.$$

The x-th row of the transition matrix P is the distribution $P(x, \cdot)$ on Ω. By the Chapman–Kolmogorov equations, we know that the t-step transition probabilities $\mathbf{P}(X_t=y \mid X_0=x)$ are just the entries in the matrix P^t.

Two important properties of Markov chains are irreducibility and periodicity. A Markov chain with transition matrix P is *irreducible* if for any two states $x, y \in \Omega$, there exists t (possibly depending on x and y) such that $P^t(x, y) > 0$. In other words, for an irreducible Markov chain, it is possible to reach any state from any other state in the state space Ω. Let $\mathcal{T}_x = \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for a Markov chain to return to its starting state x. We call the greatest common divisor of $\mathcal{T}_x$ the *period* of state x. If a Markov chain consists of only states with period 1 it is called *aperiodic*; otherwise, it is called a *periodic chain*. The assumptions of irreducibility and aperiodicity lead to the fundamental result of Markov chains, called the convergence theorem, stated below.

Theorem 2.1 [Levin et al. 2009, Theorem 4.9]. *For an irreducible and aperiodic Markov chain on a finite state space Ω with transition matrix P, there exists a unique probability distribution π on Ω, called its* stationary distribution, *such that $\pi P = \pi$ and $P^t(x, \cdot)$ converges to π as $t \to \infty$ for all initial states x.*

By the definition of the stationary distribution, if P is symmetric, then π is the uniform distribution on Ω since P is a stochastic matrix.

The mixing time is then a measure of the convergence rate of the chain to its stationary distribution, quantified in terms of a choice of a metric. As per the standard convention, we adopt the *total variation distance*: given two distributions μ and ν on a common state space Ω, the total variation distance is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \qquad (2\text{-}1)$$

Using this metric, we define the *maximal distance to stationary* of a Markov chain with transition matrix P to its stationary distribution π to be

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}}.$$

Then, given any $\varepsilon > 0$, the *mixing time* of the Markov chain is defined by

$$t_{\mathrm{mix}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}.$$

We note that it is somewhat conventional to fix a particular value of ε, and often specifically the value $\varepsilon = \frac{1}{4}$, when comparing mixing-time results. We too will adhere to this convention at times.

Finally, as we will show in Section 4, rather than obtaining bounds on $d(t)$, it is sometimes more convenient to bound the *standardized maximal distance* defined by

$$\bar{d}(t) := \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\mathrm{TV}}, \qquad (2\text{-}2)$$

which satisfies the following result.

Lemma 2.2 [Levin et al. 2009, Lemma 4.11]. *With $d(t)$ and $\bar{d}(t)$ defined above, we have*

$$d(t) \leq \bar{d}(t) \leq 2\, d(t).$$

See [Levin et al. 2009] for more on mixing times of Markov chains.

3. The rook's walk

The rook's walk describes a rook moving on an n^d-chessboard. The rook moves according to a uniformly random distribution so that any square available to it has an equal likelihood of being selected in the next move. The following definition introduces notation to formalize the intuitive notion of a legal move on the n^d-chessboard.

Definition 3.1. For integers $n \geq 3$ and $d \geq 1$, the n^d *rook's walk* is the irreducible, aperiodic, and symmetric Markov chain on the d-fold Cartesian product $\{1, \ldots, n\}^d$ with the transition probabilities

$$P(x, y) = \frac{1}{d(n-1)} \mathbf{1}_{\{\|y - x\|_0 = 1\}},$$

where $\|x\|_0 = \sum_{i=1}^d \mathbf{1}_{\{x^i \neq 0\}}$ is the Hamming distance.

By Theorem 2.1, the n^d rook's walk converges to a unique stationary distribution. Moreover, since the transition matrix for the n^d rook's walk is symmetric, the stationary distribution is the uniform distribution on the n^d possible states.

As mentioned in the introduction, one approach to finding the mixing time of this Markov chain is the spectral method, bounding the mixing time in terms of the eigenvalues of the transition matrix P. Following such a method, the bounds below for the rook's walk mixing times were obtained in [Kim 2012].

Proposition 3.2. *For the n^d rook's walk with $n \geq 3$, we have*

$$\frac{d(n-1)}{n} \log \frac{1}{2\varepsilon} \leq t_{\mathrm{mix}}(\varepsilon) \leq \frac{d(n-1)}{n} \log \frac{n^d}{\varepsilon}. \tag{3-1}$$

In [Li and Tucker 2014], by direct computation of the maximal distance to stationary $d(t)$ for the rook's walk in two dimensions, the authors derived the following result.

Proposition 3.3. *For the n^2 rook's walk with $n \geq 3$, we have*

$$t_{\mathrm{mix}}\left(\tfrac{1}{4}\right) = \begin{cases} 2 & \text{for } 3 \leq n \leq 7, \\ 3 & \text{for } n \geq 8. \end{cases}$$

That the mixing time for the rook's walk is asymptotically constant (in n) is rather intuitive: since the rook can move arbitrarily far along any row or column, increasing the length of the board does not increase the number of moves required to reach any square. This renders the length of the board to be of little consequence.

4. Coupling and path coupling methods

One of the advantages of defining the mixing time of a Markov chain in terms of the total variation distance (2-1) is that the total variation distance can be expressed in terms of couplings of distributions, which provide a powerful probabilistic tool in the analysis of mixing times.

A *coupling* of two distributions μ and ν is a pair (X, Y) of random variables defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν. The relationship between couplings and total variation distance between two distributions is given by the following proposition:

Proposition 4.1 [Levin et al. 2009, Proposition 4.7]. *Let μ and ν be two distributions on the state space Ω. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf\{\mathbf{P}(X \neq Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

Moreover, a coupling (X, Y) which attains the infimum exists and we call such a coupling the optimal coupling.

There are several applications of couplings to the analysis of probability distributions (see, e.g., [Lindvall 2002]), but we will be more interested in the extension of this idea to coupling Markov chains.

Definition 4.2. We define a *coupling of a Markov chain* with transition matrix P to be a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with common transition matrix P.

Given a coupling of a Markov chain with initial states $X_0 = x_0$ and $Y_0 = y_0$, we will write $\mathbf{P}_{x_0, y_0}$ for the probability on the product space of the process (X_t, Y_t). From Proposition 4.1, we get the following bound on the total variation distance of coupled Markov chains.

Theorem 4.3. *If (X_t, Y_t) is a coupling of a Markov chain transition matrix P and initial states with $X_0 = x_0$ and $Y_0 = y_0$, then*

$$\| P^t(x_0, \cdot) - P^t(y_0, \cdot) \|_{\mathrm{TV}} \leq \mathbf{P}_{x_0, y_0}(X_t \neq Y_t).$$

The *coupling method* for bounding mixing times of Markov chains is a process for bounding the value $\mathbf{P}_{x_0, y_0}(X_t \neq Y_t)$. There are a couple of standard techniques to bound $\mathbf{P}_{x_0, y_0}(X_t \neq Y_t)$. The approach we employ is in terms of a given a metric ρ on the state space Ω, and in our setting a natural such metric arises. Namely, given a symmetric, irreducible, and aperiodic Markov chain, we can view the elements of its state space Ω as vertices of a connected, undirected graph, with an edge between x and y precisely when $P(x, y) > 0$. Every such graph has a natural metric associated to it: for $x, y \in \Omega$, we define $\rho(x, y)$ to be the *geodesic distance* from x to y, that is, the length (in number of edges) of the shortest path from x to y in Ω. The presence of this geodesic metric ρ allows for the reformulation of several key ideas: legal moves are encoded as pairs of states with $\rho(x, y) = 1$; we can make sense of the diameter of Ω, defined by $\mathrm{diam}\, \Omega = \max_{x, y \in \Omega} \rho(x, y)$; and, since the metric only takes values that are natural numbers, the event considered in Theorem 4.3 is encoded as

$$\mathbf{P}_{x_0, y_0}(X_t \neq Y_t) = \mathbf{P}_{x_0, y_0}(\rho(X_t, Y_t) \geq 1). \tag{4-1}$$

Combining Theorem 4.3, Lemma 2.2, (4-1), and (2-2), we see that to bound mixing times it suffices to bound the probability $\mathbf{P}_{x_0, y_0}(\rho(X_t, Y_t) \geq 1)$. Generally speaking, a strong bound on this probability for arbitrary time step t is not feasible, and so we turn to the alternative method of deriving a bound of the form

$$\mathbf{E}_{x, y}[\rho(X_1, Y_1)] = \mathbf{E}\big[\rho(X_t, Y_t) \mid X_{t-1} = x, Y_{t-1} = y\big] \leq e^{-\alpha} \rho(x, y)$$

for some $\alpha > 0$. We refer to the above behavior as *contraction* of the coupled Markov chains. Then iterating over the time steps t yields a bound on the maximal distance to stationary $d(t)$. This result is stated in the following proposition.

Proposition 4.4. *Suppose* (X_t, Y_t) *is a coupling of an aperiodic and irreducible Markov chain on a finite state space* Ω, *and let* ρ *be an integer-valued metric on* Ω. *If for all* $(x, y) \in \Omega \times \Omega$,

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha} \rho(x, y)$$

for some real number $\alpha > 0$, *then*

$$t_{\mathrm{mix}}(\varepsilon) \le \left\lceil \frac{1}{\alpha} \log \frac{\mathrm{diam}\,\Omega}{\varepsilon} \right\rceil.$$

Proof. Let $x_0, y_0 \in \Omega$. By Theorem 4.3 and (4-1), we have

$$\| P^t(x_0, \cdot) - P^t(y_0, \cdot) \|_{\mathrm{TV}} \le \mathbf{P}_{x_0, y_0}(X_t \ne Y_t) = \mathbf{P}_{x_0, y_0}(\rho(X_t, Y_t) \ge 1)$$

and Markov's inequality yields

$$\mathbf{P}_{x_0, y_0}(\rho(X_t, Y_t) \ge 1) \le \mathbf{E}_{x_0, y_0}[\rho(X_t, Y_t)].$$

By the contraction assumption $\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha} \rho(x, y)$, we have

$$\mathbf{E}_{x_0, y_0}[\rho(X_t, Y_t)] = \mathbf{E}_{x_0, y_0}\big[\mathbf{E}_{x_0, y_0}[\rho(X_t, Y_t) \mid (X_{t-1}, Y_{t-1})]\big]$$
$$\le e^{-\alpha}\, \mathbf{E}_{x_0, y_0}[\rho(X_{t-1}, Y_{t-1})].$$

Repeated iteration over t time steps yields

$$\mathbf{E}_{x_0, y_0}[\rho(X_t, Y_t)] \le e^{-\alpha t} \rho(x_0, y_0).$$

Therefore, since x_0 and y_0 were arbitrary, by Lemma 2.2 we conclude that

$$d(t) \le e^{-\alpha t} \, \mathrm{diam}\,\Omega. \qquad \square$$

The above coupling method for bounding mixing times of Markov chains requires bounding the coupling distance probability $\mathbf{P}_{x,y}(\rho(X_t, Y_t) \ge 1)$ for *all* states x and y which is often nontrivial. For contrast, the *path coupling* method derived by Bubley and Dyer [1997], essentially a combination of the coupling method with the triangle inequality, requires only the demonstration of the contraction bound for *adjacent* states. The full path-coupling result is stated in the following proposition.

Proposition 4.5. *Suppose* (X_t, Y_t) *is a coupling of an irreducible and aperiodic Markov chain on a finite state space* Ω *with its geodesic metric* ρ. *If for all adjacent states* x *and* y *we have*

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha}$$

for some constant $\alpha > 0$, *then*

$$t_{\mathrm{mix}}(\varepsilon) \le \left\lceil \frac{1}{\alpha} \log \frac{\mathrm{diam}\,\Omega}{\varepsilon} \right\rceil.$$

Proof. Let $x, y \in \Omega$, let $r = \rho(x, y)$, and choose a length-r path from x to y

$$x = x_0, x_1, \ldots, x_{r-1}, x_r = y.$$

Such a path exists by the definition of ρ, and we have $\rho(x_i, x_{i-1}) = 1$ for all $i = 1, \ldots, r$. If we denote by $(X_{1,i}, X_{1,i-1})$ the coupling corresponding to (x_i, x_{i-1}), then by the hypotheses of the theorem, there is a positive constant α such that

$$\mathbf{E}_{x_i, x_{i-1}}[\rho(X_{1,i}, X_{1,i-1})] \leq e^{-\alpha}$$

for each i. Then by the triangle inequality we have

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \leq \sum_{i=1}^{r} \mathbf{E}_{x_i, x_{i-1}}[\rho(X_{1,i}, X_{1,i-1})] \leq \rho(x, y)e^{-\alpha}.$$

The result then follows from Proposition 4.4. □

In the case of rooks moving on an n^d-chessboard, the geodesic metric is described easily in terms of the motion of the rooks — the distance between two states/squares is simply the minimum number of rook moves to connect two such squares. Alternatively, viewing our chessboard as being coordinatized by the d principal axes, the distance between two states is simply the number of coordinates in which the two states' components differ (i.e., the Hamming distance between their coordinate vectors). That is, if we let $x = [x^1, \ldots, x^d]$ and $y = [y^1, \ldots, y^d]$ be squares in the d-dimensional chessboard state space Ω, then the geodesic metric ρ on Ω is given by

$$\rho(x, y) = \sum_{i=1}^{d} \mathbf{1}_{\{y^i - x^i \neq 0\}}.$$

It is worth briefly unpacking this in the language of two rooks occupying two states on the standard two-dimensional board. The case $\rho = 2$ means that the two rooks have neither a column nor row in common. Similarly $\rho = 1$ means that the two rooks share either a row or a column, and we have $\rho = 0$ if and only if two rooks occupy the same space on the board. Note that this trichotomy of metric values is precisely the trichotomy we used in the introduction to count two-move sequences.

5. Coupling of the rook's walk

The key to obtaining a good bound on the mixing time of a Markov chain using the coupling method is to construct an optimal or near-optimal coupling, i.e., one with very small values of $P(X_t \neq Y_t)$. In other words, you want to construct a coupling that encourages the coordinate processes of (X_t, Y_t) to meet as fast as possible. Below we will construct a Markov chain Y that accomplishes this for the rook's walk

on the d-dimensional chessboard Ω, and use the path coupling technique derived in the previous section to bound the mixing time. We will continue to use subscripts to denote the time step of the chain, and superscripts to denote the components of the vector describing the state; i.e., X_t^i denotes the i-th component of X at time t.

Definition 5.1 (rook's walk coupling). Let (X_t) denote the rook's walk, and (X_t, Y_t) be a process with current states $X_t = x_t$ and $Y_t = y_t$, and let $X_{t+1} = x_{t+1}$. Then for $t \geq 0$, we define Y_{t+1} using x_t, y_t, and x_{t+1}. Since $\rho(x_t, x_{t+1}) = 1$ for all values of x_{t+1}, there is a unique l such that $x_t^l \neq x_{t+1}^l$.

- If $x_{t+1}^l \neq y_t^l$, set

$$Y_{t+1}^i = \begin{cases} y_t^i & \text{for } i \neq l, \\ x_{t+1}^i & \text{for } i = l. \end{cases}$$

- Suppose $x_{t+1}^l = y_t^l$. If $x_{t+1} = y_t$, set $Y_{t+1} = x_t$. Otherwise choose uniformly one of the indices m for which $y_t^m \neq x_{t+1}^m$, and set

$$Y_{t+1}^i = \begin{cases} y_t^i & \text{for } i \neq m, \\ x_{t+1}^i & \text{for } i = m. \end{cases}$$

Let us give a more informal and intuitive explanation in the language of rooks moving on an n-dimensional chessboard. Slightly abusing terminology and notation, let X and Y denote two rooks moving on a board, with X moving according to the standard rook's walk. Then the movement of Y is to respond a move of X as directed by the following rule:

> When X moves in a given dimension, move Y along that same dimension so as to match X's component in that dimension, unless those components already match, in which case:
>
> - randomly choose another dimension for Y to move along, and match X's component in that dimension; unless
> - X moves to the square currently inhabited by Y, then lacking anything more clever to do, Y moves to X's previous location.

In short, Y will move so as to decrease $\rho(X, Y)$ by 1 whenever possible, moving along the same dimension as X unless X and Y agree in that dimension. In this case, X moves along a dimension chosen uniformly between all other axes in which their components differ. Once Y occupies the same square as X, it will move to mirror X's move every turn. In particular, note that $\rho(X, Y)$ is nonincreasing, and decreases by 0 or 1 each move until the two rooks occupy the same state.

Example 5.2. Figure 2 shows a two-dimensional walk according to this coupling for $n = 4$. The white rook moves at random via the uniform distribution on the set of legal moves, and the black rook moves deterministically via the rules of
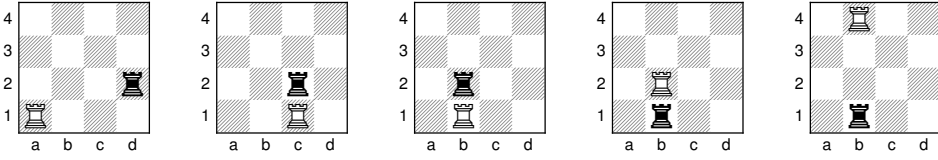
Figure 2. An instance of the coupling, with $d = 2$ and $n = 4$.

Definition 5.1. In algebraic chess notation, the move order would read 1. Rc1 Rc2 2. Rb1 Rb2 3. Rb2 Rb1 4. Rb4. The move following the last diagram will be 4. ... Rb4, at which point the two rooks' positions will coincide from then on.

The content of the following theorem can be paraphrased rather nicely in this language: if X moves according to the uniform distribution on legal moves, and Y "follows" X as prescribed by the rules above, then Y's position also satisfies the uniform distribution.

Theorem 5.3. *Let (X_t) be the rook's walk with transition matrix P. Then the process (Y_t) described above is also a Markov chain with transition matrix P; i.e., (X_t, Y_t) forms a coupling of the rook's walk Markov chain.*

Proof. The transition probabilities for X are given by the uniform distribution on legal squares, and we need to show the same is true for Y. By translational symmetry, we can without loss of generality suppose that $Y_t = [1, 1, \ldots, 1]$. For $2 \le k \le n$ and $1 \le m \le d$, let $y_{k,m}$ denote the state whose coordinates are all 1, except for a k in the m-th component. We need to show that

$$P\big(Y_{t+1} = y_{k,m} \mid Y_t = [1, 1, \ldots, 1]\big) = \frac{1}{d(n-1)}$$

for each of the $d(n-1)$ such choices, regardless of the value of X_t. Given such an X_t, let $r = \rho(X_t, Y_t)$. Relabeling the axes if needed, we can also without loss of generality assume that $X_t = [x^1, x^2, \ldots, x^r, 1, 1, \ldots, 1]$ with each $x^i > 1$.

Now, following the rules set forth in Definition 5.1:

- For $m > r$, we have $Y_{t+1} = y_{k,m}$ if and only if

$$X_{t+1} = [x^1, x^2, \ldots, x^r, 1, \ldots, k, \ldots, 1],$$

with the k occurring in the m-th slot, which occurs with probability $1/(d(n-1))$ by the transition probabilities for X.

- If $r = 1$ and $X_t = [k, 1, 1, \ldots, 1]$, then $Y_{t+1} = y_{k,1}$ if and only if $X_{t+1} = Y_t$, which occurs with probability $1/(d(n-1))$.

- If $r > 1$, $m \le r$ and $x_m \ne k$, we have $Y_{t+1} = y_{k,m}$ if and only if

$$X_{t+1} = [x^1, \ldots, x^{m-1}, k, x^{m+1}, \ldots, x^r, 1, \ldots, 1],$$

which again occurs with probability $1/(d(n-1))$.

- Finally, if $r > 1$, $m \leq r$ and $x_m = k$, then Y_{t+1} can only be $y_{k,m}$ if $l \leq r$, $l \neq m$, and $X_{t+1}^l = 1$, which occurs with probability $(r-1)/(d(n-1))$. In this case, $\rho(Y_t, X_{t+1}) = r - 1$, and so Y_{t+1} is obtained from Y_t by uniformly choosing one of the $r - 1$ components in which x_{t+1} and y_t differ for a total probability of

$$\frac{1}{r-1} \frac{r-1}{d(n-1)} = \frac{1}{d(n-1)}.\qquad \square$$

6. Mixing time of the rook's walk

Having established that the rook's walk coupling defined in Definition 5.1 does indeed provide us a coupling, we will now compute the expected contraction factor α (in the language of Proposition 4.5) for the coupling, which gives the following result for the mixing time of the rook's walk.

Theorem 6.1. *For the n^d rook's walk, we have*

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d(n-1)}{(d-1)(n-1)+1}} \right\rceil. \tag{6-1}$$

Proof. Let x and y be adjacent states. Then since $\rho(x, y) = 1$, there exists a unique component l for which $x^l \neq y^l$. Then, in the language of rooks, from Definition 5.1, $\rho(X_t, Y_t) = 0$ if and only if X makes one of the $n - 2$ moves along that axis and *not* onto y, and $\rho(X_t, Y_t) = 1$ otherwise. This gives an expected value of

$$\mathbf{E}\big[\rho(X_t, Y_t) \mid X_{t-1}=x, Y_{t-1}=y\big] = 0 \cdot \frac{n-2}{d(n-1)} + 1 \cdot \frac{d(n-1)-(n-2)}{d(n-1)}.$$

In the notation of Proposition 4.4, this says that we may take the value

$$\alpha = \log \frac{d(n-1)}{(d-1)(n-1)+1}.$$

Thus, combined with the observation that diam $\Omega = d$, we obtain the result from Proposition 4.4. $\square$

Taking $\varepsilon = \frac{1}{4}$, we have the numerical data in Table 1.
We note that:

- For fixed d, the upper bound in (3-1) grows logarithmically with n as $n \to \infty$, whereas (6-1) is asymptotically constant, approaching

$$\left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d}{(d-1)}} \right\rceil.$$

$d\downarrow$	$n \to$ 3	4	5	6	7	8	100	1000
2	8	6	5	5	4	4	4	4
3	14	10	9	9	8	8	7	7
4	21	16	14	13	12	12	10	10
5	29	21	19	18	17	16	14	14
10	72	54	48	45	43	42	36	36
100	1196	896	796	746	716	697	603	597

Table 1. Coupling bounds for $t_{\mathrm{mix}}\left(\frac{1}{4}\right)$ for various n and d.

- For fixed n, the upper bound in (3-1) grows quadratically with d as $d \to \infty$, whereas our bound

$$\left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d(n-1)}{(d-1)(n-1)+1}} \right\rceil \approx \frac{1}{1 - \frac{\log(d-1)}{\log d}}$$

grows like $d \log d$, regardless of n. (Try this as an exercise!) It is interesting to note that this value of $d \log d$ agrees with the asymptotic solution to the coupon collector's problem with d coupons. This suggests that a necessary and sufficient condition for the rook's walk to be thoroughly mixed is that our rook has moved at least once in each of the d dimensions.

As the motivating example, let us pay extra attention to generalizations of the standard 8^2-chessboard. For the case where we fix the dimension as $d = 2$, the actual values of t_{mix} (as always, for $\varepsilon = \frac{1}{4}$) are known from Proposition 3.3 for all lengths n, and so serve as a litmus test for the accuracy of a given bound. Let us denote by t_c the "unceilinged" version of the coupling bound given in Theorem 6.1, and by t_s the upper bound obtained by the spectral method in Proposition 3.2. Figure 3 demonstrates the two upper bounds t_s and t_c in conjunction with the exact values of $t_{\mathrm{mix}}\left(\frac{1}{4}\right)$ from Proposition 3.3.

In particular, it is easy to show that for $d = 2$, we have

$$\lim_{n\to\infty} t_c \le 1 - \log_2 \varepsilon,$$

and so for $\varepsilon = \frac{1}{4}$, we get an asymptotic mixing time of precisely 3, making the upper bound (6-1) asymptotically tight.

On the other hand, if we fix the length as $n = 8$ and vary instead the dimension of the board, it quickly becomes computationally difficult to evaluate exact values — we were only able to easily compute the exact mixing times for $d = 2, 3, 4$ (with respective mixing times of 3, 5, and 7). See Figure 4.

In short, it seems that the coupling method provides a stronger alternative to the standard spectral method for bounding mixing times of rook's walk Markov chains
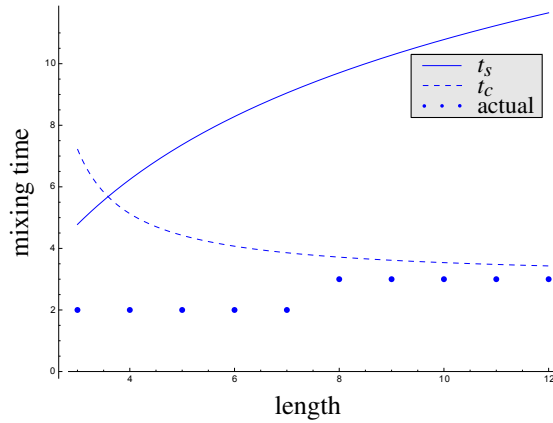
Figure 3. Bounds and actual values for $t_{\min}\left(\frac{1}{4}\right)$, $d = 2$, n varying.
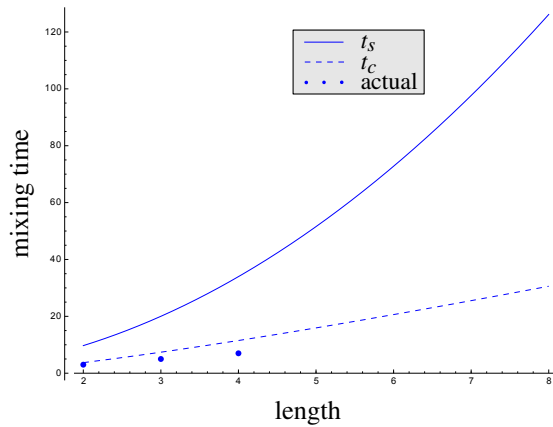
Figure 4. Bounds and actual values for $t_{\min}\left(\frac{1}{4}\right)$, $n = 8$, d varying.

from above. Further, in the two-dimensional case, the coupling bounds are asymptotically tight. There are natural extensions to the rook's walk that would be interesting to pursue. In particular, future work will address the case where the rook's moves are restricted to traversing at most k squares per move, and the case where the transition probabilities to all the allowable squares are not uniform.

Acknowledgments

References

[Bubley and Dyer 1997] R. Bubley and M. E. Dyer, "Path coupling: a technique for proving rapid mixing in Markov chains", pp. 223–231 in *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science* (Miami, FL, 1997), Institute of Electrical and Electronics Engineers, Los Alamitos, CA, 1997.

[Kemeny et al. 1976] J. G. Kemeny, J. L. Snell, and A. W. Knapp, *Denumerable Markov chains*, 2nd ed., Graduate Texts in Mathematics **40**, Springer, New York, NY, 1976. MR Zbl

[Kim 2012] S. S. Kim, "Mixing time of a rook's walk", Undergraduate certificate paper, Princeton University, 2012, available at http://www.pacm.princeton.edu/documents/Kim.pdf.

[Levin et al. 2009] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009. MR Zbl

[Li and Tucker 2014] Y. Li and K. Tucker, "Investigating the rook's walk", Senior thesis, Willamette University, Salem, OR, 2014.

[Lindvall 2002] T. Lindvall, *Lectures on the coupling method*, Dover, Mineola, NY, 2002. MR Zbl

mclemanc@umflint.edu   Department of Mathematics, University of Michigan–Flint, Flint, MI 48502 United States

potto@willamette.edu   Department of Mathematics, Willamette University, Salem, OR 97301, United States

johnr4@vt.edu   Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States

masutter@umflint.edu   Department of Mathematics, University of Michigan–Flint, Flint, MI 48502 United States

The lifting of graphs to 3-uniform hypergraphs and some applications to hypergraph Ramsey theory

Mark Budden, Josh Hiller, Joshua Lambert and Chris Sanford

(Communicated by Joshua Cooper)

Given a simple graph Γ, we describe a "lifting" to a 3-uniform hypergraph $\varphi(\Gamma)$ that sends the complement of Γ to the complement of $\varphi(\Gamma)$. We consider the effects of this lifting on cycles, complete subhypergraphs, and complete subhypergraphs missing a single hyperedge. Our results lead to natural lower bounds for some hypergraph Ramsey numbers.

1. Introduction

The subject of extremal graph theory arose from the observation that as the cardinality of a set increases, one becomes able to predict the existence of specific complex structures within the set. In particular, Ramsey theory provides a plentiful garden, ripe with open problems for all extremal graph theorists. In Ramsey theory, mathematicians focus their attention on the determination of the Ramsey number $R(K_s, K_t)$, defined to the be least natural number n with the following property: if a graph G has order at least n, then G contains a K_s-subgraph (a subgraph isomorphic to the complete graph K_s on s vertices), or the complement $\bar{G}$ contains a K_t-subgraph. While only a handful of Ramsey numbers are known, many Ramsey numbers can be found to live within certain bounds (see Radziszowski's dynamic survey [1994] for a current list of known values and restrictions). Since determining exact values for Ramsey numbers is very difficult, we often shift our attention to finding a specific graph H that does not contain a K_s-subgraph and whose complement $\bar{H}$ does not contain a K_t-subgraph to improve upon known lower bounds for these elusive numbers.

The self-complementary graphs known as Paley graphs provide a natural lower bound for the diagonal Ramsey numbers, which take the form $R(K_s, K_s)$. To define the Paley graph G_q, let

$$q = p^f \equiv 1 \pmod 4$$

be a power of the prime number p and let $\mathbb{F}_q$ denote the finite field containing exactly q elements. Then G_q has vertex set $V(G_q) := \mathbb{F}_q$ and edge set

$$E(G_q) := \{ab \mid b - a \in \mathbb{F}_q^{\times 2}\},$$

where $\mathbb{F}_q^{\times 2}$ denotes the subgroup of the multiplicative group $\mathbb{F}_q^{\times}$ consisting of squares:

$$\mathbb{F}_q^{\times 2} := \{y \in \mathbb{F}_q^{\times} \mid y = x^2 \text{ for some } x \in \mathbb{F}^{\times}\}.$$

Note that the assumed congruence $q \equiv 1 \pmod 4$ implies that $-1 \in \mathbb{F}_q^{\times 2}$ and hence, $a - b \in \mathbb{F}_q^{\times 2}$ if and only if $b - a \in \mathbb{F}_q^{\times 2}$.

Determining the aforementioned lower bound for a Ramsey number coincides with finding the clique number of a graph G, which we shall denote by $\omega(G)$ throughout this paper, along with the clique number of its complement $\bar{G}$. The clique number of a hypergraph is denoted analogously. Early discovery of lower bounds for Ramsey numbers hinged upon the results $\omega(G_5) = 2$ and $\omega(G_{17}) = 3$, which gave us $R(K_3, K_3) \geq 6$ and $R(K_4, K_4) \geq 18$. In fact, these bounds are optimal since $R(K_3, K_3) = 6$ and $R(K_4, K_4) = 18$. With the algebraic structure of Paley graphs providing a methodology for the determination of certain clique numbers, numerous generalizations of the concept of a Paley graph have been introduced (see [Budden et al. 2011; 2013] and [Su et al. 2002; Wu et al. 2010], where new lower bounds for several Ramsey numbers resulted).

One generalization of Ramsey theory worth considering is the corresponding theory in the context of 3-uniform hypergraphs. With Paley graphs playing such a vital role in the determination of the diagonal Ramsey numbers, we wish to determine the analogues of Paley graphs in this context. After some investigation, we noticed that the Paley graph G_q can be used to define an analogous hypergraph $G_q^{(3)}$ by setting $V(G_q^{(3)}) := \mathbb{F}_q$ and defining the hyperedge set

$$E(G_q^{(3)}) := \{abc \mid (b - a)(c - b)(a - c) \in \mathbb{F}_q^{\times 2}\}.$$

Then $G_q^{(3)}$ is self-complementary and maintains much of the algebraic structure inherent in Paley graphs. In fact, using a character sum similar to the one used to enumerate triangles in character difference graphs in [Budden et al. 2011; 2013], one can easily show that $G_q^{(3)}$ contains exactly

$$\tfrac{1}{192} q(q - 1)(q - 3)(q - 5)$$

subhypergraphs isomorphic to $K_4^{(3)}$ (where $K_n^{(3)}$ denotes the complete 3-uniform hypergraph on n vertices). From this calculation, we see that the first 3-uniform Paley graph that contains a $K_4^{(3)}$-subhypergraph is $G_{13}^{(3)}$, and it is well-known that $R(K_4^{(3)}, K_4^{(3)}; 3) = 13$ (see [McKay and Radziszowski 1991]). Here, $R(K_s^{(r)}, K_t^{(r)}; r)$ is the Ramsey number for r-uniform hypergraphs.

The observation that $G_q^{(3)}$ seems to be the appropriate analogue for Paley graphs in the 3-uniform setting led us to consider how an arbitrary graph might naturally be lifted to form a 3-uniform hypergraph, while maintaining properties that are useful to Ramsey theory. In Section 2, we describe a natural way to lift a graph to a 3-uniform hypergraph, show that our lifting preserves complements, and consider the lifting of cycles. In Section 3, we consider which graphs map to complete subhypergraphs and complete subhypergraphs missing a single hyperedge, allowing us to relate the clique number of a graph to that of its 3-uniform lifting.

In Section 4, we focus on applications of our results to generalized Ramsey theory. One of the more well-known results in hypergraph Ramsey theory is the "stepping-up" lemma, usually credited to Erdős and Hajnal (see [Graham et al. 1990]). It states that if $s > r \geq 3$, then

$$R(K_s^{(r)}, K_s^{(r)}; r) > m \quad \Longrightarrow \quad R(K_{2s+r-4}^{(r+1)}, K_{2s+r-4}^{(r+1)}; r+1) > 2^m.$$

Despite the strength of this result, it begins with $r = 3$, for which there exist only a small number of known lower bounds. In fact, the only known 3-uniform Ramsey number for complete hypergraphs is $R(K_4^{(3)}, K_4^{(3)}; 3) = 13$ (see [Radziszowski 1994]), but many new bounds have recently been determined for Ramsey numbers of various hypergraphs that are not complete; see [Budden et al. 2015]. A weak version of Theorem 9 in Section 4 implies that when $s \geq 3$ and $t \geq 3$, we have

$$R(K_{2s-1}^{(3)}, K_{2t-1}^{(3)}; 3) \geq R(K_s, K_t).$$

This allows one to use known lower bounds for diagonal Ramsey numbers to deduce bounds for corresponding higher-uniform Ramsey numbers via the stepping-up lemma.

2. Lifting graphs to 3-uniform hypergraphs

Let $\mathcal{G}_2$ denote the set of all (simple) graphs of order at least three and let $\mathcal{G}_3$ denote the set of all 3-uniform (simple) hypergraphs of order at least three. Define the map $\varphi : \mathcal{G}_2 \to \mathcal{G}_3$ to send a graph Γ to a 3-uniform hypergraph $\varphi(\Gamma)$ satisfying $V(\varphi(\Gamma)) = V(\Gamma)$, and letting $E(\varphi(\Gamma))$ consist of all unordered 3-tuples abc of distinct vertices in $V(\varphi(\Gamma))$ such that exactly one or all of ab, bc, and ac are in $E(\Gamma)$. We easily confirm that $\varphi(G_q) = G_q^{(3)}$, as we defined in the previous section. One can also check that if two graphs in $\mathcal{G}_2$ are isomorphic, then their images under the lifting φ must also be isomorphic. It is easily demonstrated that the converse is not true.

Denoting the complement of a graph (or hypergraph) Γ by $\overline{\Gamma}$, we note that $abc \in E(\varphi(\Gamma))$ if and only if all three of ab, bc, and ac are edges in Γ (and hence, none of them form edges in $\overline{\Gamma}$) or if exactly one of ab, bc, and ac is an edge in Γ (in

which case, exactly two of them form edges in $\overline{\Gamma}$). Observing that $\varphi(\overline{\Gamma})$ consists of hyperedges abc such that exactly zero or two of ab, bc, and ac are in Γ, it follows that

$$\overline{\varphi(\Gamma)} \cong \varphi(\overline{\Gamma}).$$

In particular, if Γ is self-complementary, then $\varphi(\Gamma)$ is self-complementary. The preservation of complements under the map φ further emphasizes this choice of lifting for its potential implications in Ramsey theory.

In order to gain an understanding of the map φ, we begin by considering its effects on cycles. For any (hyper)graph G and subset $S \subseteq V(G)$, we shall use $G[S]$ to denote the sub(hyper)graph of G induced by S. We employ the standard notation of writing

$$x_1 - x_2 - x_3 - \cdots - x_n - x_1$$

to indicate that the vertices $x_1, x_2, x_3, \ldots, x_n$ in a graph $\Gamma \in \mathcal{G}_2$ form a cycle of length n. There are two possible concepts of cycles in the 3-uniform case: loose and tight cycles. We say that $x_1, x_2, x_3, \ldots, x_n$ form a loose cycle in $\varphi(\Gamma)$ if

$$x_1 x_2 x_3, \ x_3 x_4 x_5, \ x_5 x_6 x_7, \ \ldots, \ x_{n-1} x_n x_1$$

are all hyperedges. We say that $x_1, x_2, x_3, \ldots, x_n$ form a tight cycle in $\varphi(\Gamma)$ if

$$x_1 x_2 x_3, \ x_2 x_3 x_4, \ x_3 x_4 x_5, \ \ldots, \ x_n x_1 x_2$$

are all hyperedges. Note that for loose cycles, it is necessary that n be even and every even tight cycle is also a loose cycle (having fewer hyperedges). Given a cycle $x_1 - x_2 - x_3 - \cdots - x_n - x_1$ in $\Gamma \in \mathcal{G}_2$, we first focus on when its image

$$\varphi(\Gamma)[x_1, x_2, x_3, \ldots, x_n]$$

forms a loose or tight cycle in $\varphi(\Gamma)$. The liftings of cycles when $n = 3, 4$ are easy to work out and the following two theorems handle the remaining cases.

Theorem 1. *Let $x_1 - x_2 - x_3 - \cdots - x_n - x_1$ be a cycle in $\Gamma \in \mathcal{G}_2$ with $n > 5$. If n is even and*

$$x_1 - x_3 - x_5 - \cdots - x_{n-1} - x_1$$

is a cycle in Γ, then $x_1, x_2, x_3, \ldots, x_n$ form a loose cycle in $\varphi(\Gamma)$, and if it is a cycle in $\overline{\Gamma}$, then $x_1, x_2, x_3, \ldots, x_n$ form a loose cycle in $\varphi(\overline{\Gamma})$.

Proof. Assuming that $x_1 - x_2 - x_3 - \cdots - x_n - x_1$ is a cycle in Γ, it follows that for each potential hyperedge $x_{i-1} x_i x_{i+1}$, both $x_{i-1} x_i$ and $x_i x_{i+1}$ form edges in Γ. Thus, $x_{i-1} x_i x_{i+1}$ is a hyperedge in $\varphi(\Gamma)$ if and only if $x_{i-1} x_{i+1}$ is an edge in Γ and it is a hyperedge in $\varphi(\overline{\Gamma})$ if and only if $x_{i-1} x_{i+1}$ is an edge in $\overline{\Gamma}$. $\square$
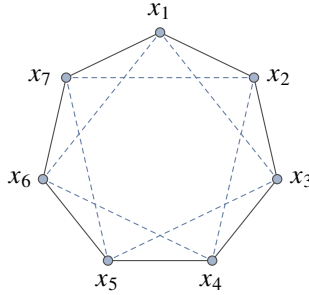
Figure 1. Parallel cycles when $n = 7$.

Theorem 2. *Let $x_1 - x_2 - x_3 - \cdots - x_n - x_1$ be a cycle in $\Gamma \in \mathcal{G}_2$ with $n \geq 5$. If n is odd and*

$$x_1 - x_3 - x_5 - \cdots - x_n - x_2 - x_4 - x_6 - \cdots - x_{n-1} - x_1$$

is a cycle in Γ, then $x_1, x_2, x_3, \ldots, x_n$ form a tight cycle in $\varphi(\Gamma)$, and if it is a cycle in $\overline{\Gamma}$, then $x_1, x_2, x_3, \ldots, x_n$ form a tight cycle in $\varphi(\overline{\Gamma})$. If n is even and both

$$x_1 - x_3 - x_5 - \cdots - x_{n-1} - x_1 \quad and \quad x_2 - x_4 - x_6 - \cdots - x_n - x_2$$

form cycles in Γ, then $x_1, x_2, x_3, \ldots, x_n$ form a tight cycle in $\varphi(\Gamma)$, and if they are both cycles in $\overline{\Gamma}$, then $x_1, x_2, x_3, \ldots, x_n$ form a tight cycle in $\varphi(\overline{\Gamma})$.

Proof. This theorem follows from a similar argument to the one that was used in the proof of the previous theorem. The details are left to the reader. □

Figures 1 and 2 provide visual representations for the underlying cycles in Theorem 2 when $n = 7$ and $n = 16$, respectively. In each graph, the cycle

$$x_1 - x_2 - x_3 - \cdots - x_n - x_1$$
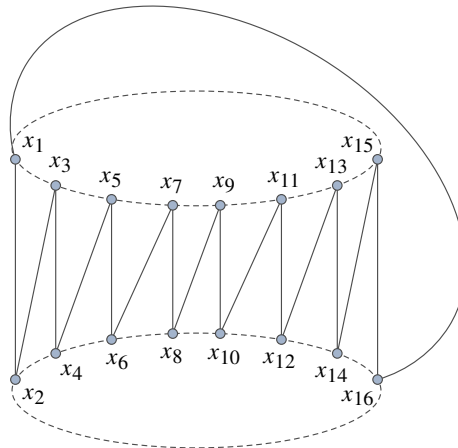
Figure 2. Parallel cycles when $n = 16$.

uses solid edges and all other cycles are represented with dashed edges. Whether or not dashed edges appear in Γ or $\overline{\Gamma}$ determines the location of the corresponding tight cycles in the lifting.

Although we are able to understand how cycles may lift, the subgraphs which map to loose and tight hypergraph cycles have a much less predictable structure, preventing such a nice characterization. So, we turn our attention to complete subhypergraphs and those that are missing a single hyperedge.

3. Complete hypergraphs and complete hypergraphs missing a single hyperedge

From the definition of φ, it is clear that if H is a complete subgraph of $\Gamma \in \mathcal{G}_2$ having order at least three, then its image $\varphi(H)$ is a complete subhypergraph of the same order in $\varphi(\Gamma)$. Now we turn our attention to understanding which subgraphs map to complete subhypergraphs under φ.

Lemma 3. *Suppose that $\Gamma \in \mathcal{G}_2$, $S \subseteq V(\Gamma)$ is a subset containing at least three elements, and $K := \Gamma[S]$. If $\varphi(K)$ is complete and C is a component of K, then C is complete.*

Proof. Suppose that C is a component of K that is not complete (which necessarily requires the order of C to be at least two). Then there exist vertices $b_1, b_2 \in V(C)$ that do not form an edge in C. If $x \in V(K) - V(C)$, then neither $b_1 x$ nor $b_2 x$ form edges in K, and $b_1 b_2 x$ is not a hyperedge in $\varphi(K)$, contradicting the assumption that $\varphi(K)$ is complete. If no such x exists, then $V(K) = V(C)$, and for every vertex $y \in V(C) - \{b_1, b_2\}$, exactly one of $b_1 y$ and $b_2 y$ must be in $E(K)$. Since C is assumed to be connected, it must have order at least four. Let N_{b_1} and N_{b_2} denote the sets of neighbors of b_1 and b_2, respectively, in $V(C)$. Note that each N_{b_j} is nonempty or else b_j would be disconnected from the rest of C. Also, since C is connected, it follows that $N_{b_1} \cap N_{b_2} \neq \varnothing$. So, let $z \in N_{b_1} \cap N_{b_2}$. Then $b_1 b_2 z$ is not a hyperedge in $\varphi(K)$, contradicting the assumption that $\varphi(K)$ is complete. Thus, we find that C must be complete. $\square$

Lemma 3 greatly restricts the structure of the possible subgraphs of a graph Γ that can map to a complete subhypergraph of $\varphi(\Gamma)$. The following theorem completely classifies the relevant subgraphs.

Theorem 4. *Suppose that $\Gamma \in \mathcal{G}_2$, $S \subseteq V(\Gamma)$ is a subset containing at least three elements, and $K := \Gamma[S]$. Then $\varphi(K)$ is complete if and only if K is complete or K is the union of exactly two disjoint complete subgraphs.*

Proof. From Lemma 3, it suffices to prove that $\varphi(K)$ is complete if and only if K contains at most two components. To prove the forward implication, assume that $\varphi(K)$ is complete and K consists of at least three components. Suppose that

C_1, C_2, C_3 are three components of K and for each $1 \le i \le 3$, choose a vertex $a_i \in V(C_i)$. Since the components are disconnected from one another, it follows that $\varphi(K)$ lacks the hyperedge $a_1 a_2 a_3$, contradicting our assumption that $\varphi(K)$ is complete. Hence, K contains at most two components. Now we consider the converse. Clearly, if K is a complete subgraph of order at least three, then $\varphi(K)$ must also be complete. Otherwise, assume that K is the disjoint union of two complete subgraphs having vertex sets S_1 and S_2. For every three vertices $a, b, c \in V(K)$, either all three are in one of S_1 or S_2, and hence form a hyperedge in $\varphi(K)$, or they are divided up between S_1 and S_2. Without loss of generality, assume that $a \in S_1$ and $b, c \in S_2$. Exactly one of ab, bc, and ac are edges in K, making $abc \in E(\varphi(K))$. It follows that $\varphi(K)$ is complete. $\qquad\square$

The previous theorem gives us an immediate corollary pertaining to the lifting of a complete bipartite graph.

Corollary 5. *For the complete bipartite graph $K_{m,n}$ where either m or n is greater than 2, we have $\varphi(K_{m,n})$ is isomorphic to the empty 3-uniform hypergraph of order $m + n$.*

Proof. Recall that $\overline{K_{m,n}} \simeq K_m \cup K_n$. Since the lifting φ preserves complements, we just apply the previous theorem to obtain our desired result. $\qquad\square$

Since every complete subgraph with at least three vertices in $\Gamma \in \mathcal{G}_2$ maps to a complete subhypergraph of $\varphi(\Gamma)$, we have

$$\omega(\varphi(\Gamma)) = m \ge 3 \quad \Longrightarrow \quad \omega(\Gamma) \le m,$$

and the previous theorem implies that

$$\omega(\Gamma) = n \ge 3 \quad \Longrightarrow \quad \omega(\varphi(\Gamma)) \le 2n.$$

From these observations, we obtain the following corollary.

Corollary 6. *Every graph $\Gamma \in \mathcal{G}_2$ with $\omega(\Gamma) \ge 3$ satisfies*

$$\omega(\Gamma) \le \omega(\varphi(\Gamma)) \le 2\omega(\Gamma),$$
$$\tfrac{1}{2}\omega(\varphi(\Gamma)) \le \omega(\Gamma) \le \omega(\varphi(\Gamma)).$$

Now let H be a subgraph of $\Gamma \in \mathcal{G}_2$ of order at least three that is isomorphic to a complete graph with a single edge missing. Without loss of generality, assume that $V(H) = \{a_1, a_2, \ldots, a_k\}$ with $a_1 a_2 \notin E(H)$. Then for every $x \in \{a_3, \ldots, a_k\}$, we have $a_1 a_2 x \notin E(\varphi(H))$. However, every unordered 3-tuple of distinct elements in $\{a_3, \ldots, a_k\}$ forms a hyperedge in $\varphi(H)$ as does any 3-tuple of vertices from $V(H)$ that contains exactly one of a_1 and a_2. So, $\varphi(H)$ is isomorphic to a complete 3-uniform hypergraph with exactly $k - 2$ hyperedges missing (those containing

a_1 and a_2). Now we consider which graphs (if any) lift under φ to hypergraphs isomorphic to complete hypergraphs missing a single hyperedge.

Theorem 7. *Suppose $n \geq 4$ and $\Gamma \in \mathcal{G}_2$. The lifting $\varphi(\Gamma)$ cannot contain an induced subhypergraph isomorphic to $K_n^{(3)} - e$ (i.e., a complete 3-uniform hypergraph on n vertices that is missing a single hyperedge).*

Proof. Assume $\varphi(\Gamma)$ contains an induced subhypergraph isomorphic to $K_n^{(3)} - e$. Let $S = \{b_1, b_2, \ldots, b_n\}$ denote the vertices of the $(K_n^{(3)} - e)$-subhypergraph of $\varphi(\Gamma)$. Without loss of generality, let $b_1 b_2 b_3$ be the missing hyperedge. There exist two possibilities. Either none of $b_1 b_2$, $b_2 b_3$, and $b_1 b_3$ are in $E(\Gamma)$ or exactly two of the aforementioned edges exist in Γ. In the former case, note that $b_1 b_2 b_4 \in E(\varphi(\Gamma))$, from which we see that exactly one of $b_1 b_4$ and $b_2 b_4$ must be in $E(\Gamma)$. Without loss of generality, assume that $b_1 b_4 \in E(\Gamma)$. In a similar manner, it can be shown that $b_2 b_3 b_4 \in E(\varphi(\Gamma))$ implies that $b_3 b_4 \in E(\Gamma)$. Then $b_1 b_3 b_4$ cannot be contained in $E(\varphi(\Gamma))$, contradicting the assumption that $b_1 b_2 b_3$ was the only missing hyperedge. In the latter case, exactly two of $b_1 b_2$, $b_2 b_3$, and $b_1 b_3$ are in $E(\Gamma)$. Without loss of generality, assume that $b_1 b_2$ and $b_2 b_3$ are in $E(\Gamma)$. Then $b_1 b_3 b_4 \in E(\varphi(\Gamma))$ implies that exactly one of $b_1 b_4$ and $b_3 b_4$ is in $E(\Gamma)$. Without loss of generality, assume that $b_1 b_4 \in E(\Gamma)$. Then $b_1 b_2 b_4 \in E(\varphi(\Gamma))$ implies that $b_2 b_4 \in E(\Gamma)$. Similarly, $b_2 b_3 b_4 \in E(\varphi(\Gamma))$ implies that $b_3 b_4 \in E(\Gamma)$, contradicting our assumption. Hence, we have shown in both cases that if $\varphi(\Gamma)$ contains a $(K_n^{(3)} - e)$-subhypergraph, then it must contain a $K_n^{(3)}$-subhypergraph. $\qquad\square$

4. Applications to Ramsey theory

From the specific subhypergraphs that we have chosen to consider under the lifting φ, it should be obvious that our interests lie in applications to extremal graph theory. In particular, our focus on the behavior of complete sub(hyper)graphs indicates an underlying interest in Ramsey theory. The multicolor Ramsey number $R(G_1, G_2, \ldots, G_k)$ is defined to be the least natural number n such that for every arbitrary coloring of the edges of K_n with k colors, there exists a subgraph in color i isomorphic to G_i for some i. The multicolor 3-uniform hypergraph Ramsey number $R(G_1, G_2, \ldots, G_k; 3)$ is defined analogously.

When studying the behavior of cliques in graphs, a class of graphs known as Turán graphs possess certain optimal parameters. Suppose $n \geq 3$ and $q \geq 2$ are integers. By the division algorithm, there exist unique integers $m \geq 0$ and $0 \leq r < q$ such that $n = mq + r$. The Turán graph $T_q(n)$ is the complete q-partite graph whose vertices are partitioned into balanced sets (i.e., sets with cardinalities as equal as possible). Such graphs contain K_q-subgraphs but lack K_{q+1}-subgraphs. In fact, Turán [1941] proved that out of all graphs of order n, they possess the maximum

number of edges possible without containing a K_{q+1}-subgraph. When considering the lifting of Turán graphs, we obtain the following theorem.

Theorem 8. *Let $n \geq 3$, $q \geq 2$, and $n = mq + r$, where $0 \leq r < q$. Then we have the following*:

(1) *If $n = qm$, then $R(K_{q+1}^{(3)} - e, K_{2m+1}^{(3)} - e; 3) > n$.*

(2) *If $n = qm + 1$, then $R(K_{q+1}^{(3)} - e, K_{2m+2}^{(3)} - e; 3) > n$.*

(3) *If $n = qm + r$ with $r \geq 2$, then $R(K_{q+1}^{(3)} - e, K_{2m+3}^{(3)} - e; 3) > n$.*

Proof. Regardless of the value of r, note that $T_q(n)$ contains a K_q-subgraph, but not a K_{q+1}-subgraph. Also, at most one vertex of a complete subgraph can come from any one connected set of vertices. So, $\varphi(T_q(n))$ contains a $K_q^{(3)}$-subhypergraph, but not a $K_{q+1}^{(3)}$-subhypergraph. Note that $\overline{T_q(n)}$ consists of disconnected complete subgraphs of orders m and $m+1$. By Theorem 4, we obtain the following cases. If $n = qm$, then all of the sets of vertices have cardinality m and $\varphi(\overline{T_q(n)})$ contains a $K_{2m}^{(3)}$-subhypergraph, but not a $K_{2m+1}^{(3)}$-subhypergraph. If $n = qm + 1$, then exactly one vertex set has cardinality $m+1$ and $\varphi(\overline{T_q(n)})$ contains a $K_{2m+1}^{(3)}$-subhypergraph, but not a $K_{2m+2}^{(3)}$-subhypergraph. For the remaining cases, in which $n = qm + r$ with $2 \leq r < q$, at least two vertex sets have cardinality $m+1$, and we find that $\varphi(\overline{T_q(n)})$ contains a $K_{2m+2}^{(3)}$-subhypergraph, but not a $K_{2m+3}^{(3)}$-subhypergraph. These results, along with the implication of Theorem 7, prove the theorem. $\square$

Now we shift our attention to proving a relationship between standard Ramsey numbers and certain corresponding 3-uniform Ramsey numbers for complete hypergraphs missing a single hyperedge.

Theorem 9. *Let $s, t \in \mathbb{N}$ with $s \geq 3$ and $t \geq 3$. Then*

$$R(K_{2s-1}^{(3)} - e, K_{2t-1}^{(3)} - e; 3) \geq R(K_s, K_t).$$

Proof. Assume $m = R(K_s, K_t)$. Then there exists a graph G of order $m - 1$ that does not contain a K_s-subgraph, and whose complement does not contain a K_t-subgraph. From Theorem 4, it follows that $\varphi(G)$ does not contain a $K_{2s-1}^{(3)}$-subhypergraph, and its complement does not contain a $K_{2t-1}^{(3)}$-subhypergraph. Theorem 7 then implies that $\varphi(G)$ does not contain a $(K_{2s-1}^{(3)} - e)$-subhypergraph, and its complement does not contain a $(K_{2t-1}^{(3)} - e)$-subhypergraph. Thus,

$$R(K_{2s-1}^{(3)} - e, K_{2t-1}^{(3)} - e; 3) > m - 1 = R(K_s, K_t) - 1,$$

completing the proof of the theorem. $\square$

Note that Theorem 9 implies

$$R(K_{2s-1}^{(3)}, K_{2t-1}^{(3)}; 3) \geq R(K_s, K_t),$$

which can be used with the stepping-up lemma. Recently, Conlon, Fox, and Sudakov [Conlon et al. 2013] also proved an analogue of the stepping-up lemma, which lifts from graphs to 3-uniform hypergraphs. In the spirit of the original stepping-up lemma, it focused on the diagonal case. Namely, they proved that

$$R(K_s, K_s) > m \implies R(K^{(3)}_{s+1}, K^{(3)}_{s+1}, K^{(3)}_{s+1}, K^{(3)}_{s+1}; 3) > 2^m.$$

Of course since $R(4, 4) = 18$, this result implies $R(5, 5, 5, 5; 3) > 131,072$. The following theorem handles some off-diagonal cases.

Theorem 10. *If $q \geq 3$, then*

$$R(K^{(3)}_5, K^{(3)}_{q+1} - e, K^{(3)}_{2s-1} - e, K^{(3)}_{2t-1} - e; 3) > q(R(K_s, K_t) - 1).$$

Proof. Suppose $m = R(K_s, K_t)$ and $q \geq 3$, and let $n = q(m - 1)$. Denote the partitioned vertex sets in $T_q(n)$ by $V_1, V_2, \ldots, V_k$. We have already noted that $\varphi(T_q(n))$ contains a $K^{(3)}_q$-subhypergraph, but not a $K^{(3)}_{q+1}$-subhypergraph. From Theorem 7, it follows that it does not contain a $(K^{(3)}_{q+1} - e)$-subhypergraph. Color the hyperedges in $\varphi(T_q(n))$ yellow. Note that $\overline{T_q(n)}$ consists of q disconnected K_{m-1}-subgraphs. Since $R(s, t) = m$, there exists a red/blue coloring of the edges of K_{m-1} that does not contain a red K_s-subgraph or a blue K_t-subgraph. When lifting just a single K_{m-1} colored in this way, the lifted hypergraph contains at most a red $K^{(3)}_{2s-2}$-subhypergraph or a blue $K^{(3)}_{2t-2}$-subhypergraph by Theorem 4. In fact by Theorem 9, the lifted hypergraph does not contain a red $(K^{(3)}_{2s-1} - e)$-subhypergraph or a blue $(K^{(3)}_{2t-1} - e)$-subhypergraph. We apply this coloring to the hyperedges in $\varphi(\overline{T_q(n)})$ that arise from the individual liftings of the disjoint vertex sets. The remaining hyperedges in $\varphi(\overline{T_q(n)})$ are precisely those that include one vertex from V_i and the other two vertices from V_j, where $i \neq j$. Color these hyperedges green. A complete subhypergraph formed using only these edges includes at most two vertices from any V_i and vertices from no more than two of the partitioned vertex sets. Hence, the green hyperedges may contain a $K^{(3)}_4$-subhypergraph, but not a $K^{(3)}_5$-subhypergraph. From this coloring, we find that

$$R(K^{(3)}_5, K^{(3)}_{q+1} - e, K^{(3)}_{2s-1} - e, K^{(3)}_{2t-1} - e; 3) > n = q(m - 1),$$

from which the theorem follows. $\square$

Although the result of [Conlon et al. 2013] is stronger than Theorem 10 for diagonal Ramsey numbers, our results improve on many known lower bounds for off-diagonal 4-color 3-uniform Ramsey numbers. For example, using the explicit known lower bounds in Table IIc of [Radziszowski 1994], we obtain the following bound on an off-diagonal Ramsey number:

$$R(22, 22) > 29,940 \implies R(K^{(3)}_5, K^{(3)}_{43} - e, K^{(3)}_{43} - e, K^{(3)}_{43} - e; 3) > 1,257,480.$$

The main advantage to considering the lifting φ is that one is able to sufficiently restrict the structure of hypergraphs in the image by knowing the structure of graphs in the domain. Many open questions naturally arise from this construction. One obvious question is whether or not analogous liftings can be constructed from graphs to r-uniform hypergraphs. This question was recently considered in [Budden and Rapp 2015], but since the liftings did not preserve complements when $r > 3$, it did not lead to new implications in Ramsey theory. We conclude with a list of several other avenues of potential inquiry:

(1) Besides cycles, complete hypergraphs, and complete hypergraphs missing a single hyperedge, what other hypergraph images have predictable preimages?

(2) Can one classify the hypergraphs in $\mathcal{G}_3$ that are not in the range of φ?

(3) Is it possible to classify all graphs in the preimage of a particular hypergraph in the range of φ?

(4) The fact that the lifting φ preserves complements means that it can be thought of as mapping a 2-coloring of the edges of K_n to a 2-coloring of the hyperedges in $K_n^{(3)}$. Can φ be used to describe a mapping of a k-coloring of the edges in K_n to a k-coloring of the hyperedges in $K_n^{(3)}$? If so, one may be able to use known bounds for multicolor Ramsey numbers to obtain analogous results in the setting of 3-uniform hypergraphs.

References

[Budden and Rapp 2015] M. Budden and A. Rapp, "Constructing r-uniform hypergraphs with restricted clique numbers", *North Carolina J. Math. Stat.* **1** (2015), 30–34.

[Budden et al. 2011] M. Budden, N. Calkins, W. N. Hack, J. Lambert, and K. Thompson, "Enumeration of triangles in quartic residue graphs", *Integers* **11** (2011), Article ID #A48. MR Zbl

[Budden et al. 2013] M. Budden, N. Calkins, W. N. Hack, J. Lambert, and K. Thompson, "Dirichlet character difference graphs", *Acta Math. Univ. Comenian. (N.S.)* **82**:1 (2013), 21–28. MR Zbl

[Budden et al. 2015] M. Budden, J. Hiller, and A. Rapp, "Generalized Ramsey theorems for r-uniform hypergraphs", *Australas. J. Combin.* **63** (2015), 142–152. MR Zbl

[Conlon et al. 2013] D. Conlon, J. Fox, and B. Sudakov, "An improved bound for the stepping-up lemma", *Discrete Appl. Math.* **161**:9 (2013), 1191–1196. MR Zbl

[Graham et al. 1990] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, 2nd ed., Wiley, New York, NY, 1990. MR Zbl

[McKay and Radziszowski 1991] B. D. McKay and S. P. Radziszowski, "The first classical Ramsey number for hypergraphs is computed", pp. 304–308 in *Proceedings of the Second Annual ACM-SIAM Symposium on Discrete Algorithms* (San Francisco, CA, 1991), edited by A. Aggarwal, ACM, New York, NY, 1991. MR Zbl

[Radziszowski 1994] S. P. Radziszowski, "Small Ramsey numbers", *Electron. J. Combin.* **1** (1994), Dynamic Survey 1. MR Zbl

[Su et al. 2002] W. Su, Q. Li, H. Luo, and G. Li, "Lower bounds of Ramsey numbers based on cubic residues", *Discrete Math.* **250**:1–3 (2002), 197–209. MR Zbl

[Turán 1941] P. Turán, "On an extremal problem in graph theory", *Mat. Fiz. Lapok* **48** (1941), 436–452. In Hungarian. MR Zbl

[Wu et al. 2010] K. Wu, W. Su, H. Luo, and X. Xu, "A generalization of generalized Paley graphs and new lower bounds for $R(3, q)$", *Electron. J. Combin.* **17**:1 (2010), Note #N25. MR Zbl

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

jphiller1@ufl.edu   *Department of Mathematics, PO Box 118105, University of Florida, Gainesville, FL 32611, United States*

joshua.lambert@armstrong.edu   *Department of Mathematics, Armstrong Atlantic State University, 11935 Abercorn Street, Savannah, GA 31419, United States*

clsanfor@syr.edu   *Department of Mathematics, Syracuse University, Syracuse, NY 13244, United States*

msp

The multiplicity of solutions for
a system of second-order differential equations

Olivia Bennett, Daniel Brumley, Britney Hopkins,
Kristi Karber and Thomas Milligan

(Communicated by John Baxley)

Making use of the Guo–Krasnosel'skiĭ fixed point theorem multiple times, we establish the existence of at least three positive solutions for the system of second-order differential equations $-u''(t) = g\big(t, u(t), u'(t), v(t), v'(t)\big)$ and $-v''(t) = \lambda f\big(t, u(t), u'(t), v(t), v'(t)\big)$ for $t \in (0, 1)$ with right focal boundary conditions $u(0) = v(0) = 0$, $u'(1) = a$, and $v'(1) = b$, where $f, g : [0, 1] \times [0, \infty)^4 \to [0, \infty)$ are continuous, $a, b, \lambda \geq 0$, and $a + b > 0$. Our technique involves transforming the system of differential equations to a new system with homogeneous boundary conditions prior to applying the aforementioned fixed point theorem.

1. Introduction

Showing the existence of multiple positive solutions for boundary value problems is an active field of study due to the applications that arise in modeling real world phenomena. A classic example based on beam analysis, presented by Agarwal [1989], gives an existence and uniqueness result of the fourth-order problem $x^{(4)} = f(t, x, x', x'', x^{(3)})$. Additionally, do Ó, Lorca, and Ubilla [do Ó et al. 2008] studied the fourth-order nonhomogeneous boundary value problem,

$$u^{(4)} = \lambda h(t, u, u''), \quad t \in (0, 1),$$
$$u(0) = u''(0) = 0,$$
$$u(1) = a, \quad u''(1) = b.$$

Utilizing a technique of rewriting the fourth-order problem as a system of second-order differential equations, the authors guaranteed existence of multiple positive solutions by ultimately applying the Guo–Krasnosel'skiĭ fixed point theorem [Krasnosel'skiĭ 1964]. Hopkins [2015] extended this process to establish multiple solutions to the differential equation $u^{(2n)} = \lambda h(t, u, u'', \ldots, u^{2(n-1)})$ satisfying

right focal boundary conditions. Henderson and Hopkins [2010] applied this same technique to a similar fourth-order difference equation. In this work, we consider the system of second-order differential equations

$$-u''(t) = g\big(t, u(t), u'(t), v(t), v'(t)\big), \tag{1}$$

$$-v''(t) = \lambda f\big(t, u(t), u'(t), v(t), v'(t)\big), \tag{2}$$

$$u(0) = v(0) = 0, \tag{3}$$

$$u'(1) = a, \quad v'(1) = b, \tag{4}$$

where $f, g : [0, 1] \times [0, \infty)^4 \to [0, \infty)$ are continuous, $\lambda, a, b \geq 0$ and $a + b > 0$. The novelty of our paper is that the functions f and g contain both even- and odd-order derivatives.

In Section 2 of this paper, we consider a transformation of (1)–(4) that satisfies homogeneous boundary conditions. We also introduce some preliminaries and the conditions under which we can eventually apply the Guo–Krasnosel'skiĭ fixed point theorem. In Section 3 we introduce and prove a sequence of lemmas giving bounds on a defined operator. This culminates in the main result, given in Section 4 where we apply the Guo–Krasnosel'skiĭ fixed point theorem multiple times, yielding at least three positive solutions.

2. Preliminaries

We will prove the existence of multiple solutions for the system of second-order differential equations (1)–(4) by applying the transformation $\bar{u}(t) = u(t) - at$ and $\bar{v}(t) = v(t) - bt$, which gives

$$-\bar{u}''(t) = g\big(t, \bar{u}(t) + ta, \bar{u}'(t) + a, \bar{v}(t) + tb, \bar{v}'(t) + b\big), \tag{5}$$

$$-\bar{v}''(t) = \lambda f\big(t, \bar{u}(t) + ta, \bar{u}'(t) + a, \bar{v}(t) + tb, \bar{v}'(t) + b\big), \tag{6}$$

$$\bar{u}(0) = \bar{v}(0) = 0, \tag{7}$$

$$\bar{u}'(1) = 0, \quad \bar{v}'(1) = 0, \tag{8}$$

where $a, b, \lambda \geq 0$ and $a + b > 0$. Notice that solutions to (5)–(8) are in one-to-one correspondence with (1)–(4). Furthermore, suppose the following hypotheses on f and g are satisfied.

(H0) The functions $f, g : [0, 1] \times [0, \infty)^4 \to [0, \infty)$ are continuous and are nondecreasing in the second and fourth variables and nonincreasing in the third and fifth variables.

(H1) There exist $\alpha, \beta \in (0, 1)$, $\alpha < \beta$, such that given $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 \neq 0$, there exists $k > 0$ such that for $t \in [\alpha, \beta]$,

$$f(t, x_1, x_2, x_3, x_4) > k.$$

(H2) For $t \in (0, 1)$,

$$\lim_{x_1+x_2+x_3+x_4 \to 0^+} \frac{f(t, x_1, x_2, x_3, x_4)}{x_1 + x_2 + x_3 + x_4} = 0$$

uniformly.

(H3) For $t \in (0, 1)$,

$$\lim_{x_1+x_2+x_3+x_4 \to \infty} \frac{f(t, x_1, x_2, x_3, x_4)}{x_1 + x_2 + x_3 + x_4} = 0$$

uniformly.

(H4) There exist $\gamma \in \left(0, \frac{2}{3}\right)$ and $q > 0$ such that for $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 < q$,

$$g(t, x_1, x_2, x_3, x_4) \le \gamma(x_1 + x_2 + x_3 + x_4) \quad \text{for } t \in [0, 1].$$

(H5) There exist $\eta \in \left(0, \frac{2}{3}\right)$ and $\hat{\rho} > 0$ such that for $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 > \hat{\rho}$,

$$g(t, x_1, x_2, x_3, x_4) \le \eta(x_1 + x_2 + x_3 + x_4) \quad \text{for } t \in [0, 1].$$

Solutions to (5)–(8), provided they exist, are of the form

$$\bar{u}(t) = \int_0^1 G(t, s) g\big(s, \bar{u}(s) + as, \bar{u}'(s) + a, \bar{v}(s) + bs, \bar{v}'(s) + b\big) ds, \tag{9}$$

$$\bar{v}(t) = \lambda \int_0^1 G(t, s) f\big(s, \bar{u}(s) + as, \bar{u}'(s) + a, \bar{v}(s) + bs, \bar{v}'(s) + b\big) ds, \tag{10}$$

where $G(t, s)$ is the Green's function

$$G(t, s) = \begin{cases} t & \text{if } 0 \le t \le s \le 1, \\ s & \text{if } 0 \le s \le t \le 1. \end{cases}$$

Since $G(t, s)$ is clearly nonnegative and f and g are nonnegative by assumption, it follows that solutions u and v are also nonnegative. Some other useful properties on $G(t, s)$ are that

$$\max_{t \in [0,1]} \int_0^1 G(t, s) \, ds = \frac{1}{2} \quad \text{and} \quad \max_{t \in [0,1]} \int_0^1 \left| \frac{\partial}{\partial t} G(t, s) \right| ds = 1.$$

In order to make use of the Guo–Krasnosel'skiĭ fixed point theorem, we will need a Banach space and a cone, as well as an operator T. Let $(X, \|\cdot\|)$ denote the Banach space $X = C^1([0, 1], \mathbb{R}) \times C^1([0, 1], \mathbb{R})$ endowed with the norm

$$\|(\bar{u}, \bar{v})\| = \|\bar{u}\|_\infty + \|\bar{u}'\|_\infty + \|\bar{v}\|_\infty + \|\bar{v}'\|_\infty,$$

where $\|\bar{u}\|_\infty = \sup_{t \in [0,1]} |\bar{u}(t)|$.

Recall that a cone, C, in X is a nonempty, closed, convex subset of X satisfying:

(1) If $x \in C$, and $\lambda > 0$, then $\lambda x \in C$.

(2) If $x \in C$ and $-x \in C$, then $x = 0$.

Define $C \subset X$ to be the cone

$$C = \left\{ (\bar{u}, \bar{v}) \in X : (\bar{u}, \bar{v})(0) = (\bar{u}', \bar{v}')(1) = (0, 0) \text{ and } \bar{u}, \bar{v} \text{ are concave} \right\}.$$

The fact that C is a cone follows directly from the definition. Moreover, let Ω_p denote the open set $\Omega_p = \{(\bar{u}, \bar{v}) \in X : \|(\bar{u}, \bar{v})\| < p\}$. Finally, define $T : X \to X$ to be the operator $T(\bar{u}, \bar{v}) = (A_1(\bar{u}, \bar{v}), A_2(\bar{u}, \bar{v}))$, where

$$A_1 = \int_0^1 G(t, s) g\left(s, \bar{u}(s) + as, \bar{u}'(s) + a, \bar{v}(s) + bs, \bar{v}'(s) + b\right) ds$$

and

$$A_2 = \lambda \int_0^1 G(t, s) f\left(s, \bar{u}(s) + as, \bar{u}'(s) + a, \bar{v}(s) + bs, \bar{v}'(s) + b\right) ds.$$

Consider the following lemma, which provides a useful property of T.

Lemma 2.1. *The operator $T : C \to C$ is completely continuous.*

We note that one can use a standard Arzelà–Ascoli argument to show that T is completely continuous; see [Hopkins 2009].
In the next section, we will take advantage of the following lemma.

Lemma 2.2. *Let $\bar{u}(t)$ be a nonnegative concave function which is continuous on $[0, 1]$. Then for all $\alpha, \beta \in (0, 1)$, with $\alpha < \beta$, we have*

$$\inf_{t \in [\alpha, \beta]} \bar{u}(t) \geq \alpha(1 - \beta) \|\bar{u}\|_\infty.$$

For a proof of Lemma 2.2, see [Hopkins 2009].

Since we will be using the Guo–Krasnosel'skiĭ fixed point theorem multiple times to acquire our main result, we end the section with the statement of this theorem.

Theorem 2.3 (Guo–Krasnosel'skiĭ fixed point theorem). *Let $(X, \|\cdot\|)$ be a Banach space and $C \subset X$ be a cone. Suppose Ω_1, Ω_2 are open subsets of X satisfying $0 \in \Omega_1 \subset \overline{\Omega}_1 \subset \Omega_2$. If $T : C \cap (\overline{\Omega}_2 \setminus \Omega_1) \to C$ is a completely continuous operator such that either*

(1) $\|Tu\| \leq \|u\|$ *for $u \in C \cap \partial\Omega_1$ and $\|Tu\| \geq \|u\|$ for $u \in C \cap \partial\Omega_2$, or*

(2) $\|Tu\| \geq \|u\|$ *for $u \in C \cap \partial\Omega_1$ and $\|Tu\| \leq \|u\|$ for $u \in C \cap \partial\Omega_2$,*

then T has a fixed point in $C \cap (\overline{\Omega}_2 \setminus \Omega_1)$.

3. Technical results

In this section we give a sequence of four lemmas that allow us to obtain the estimates needed to apply the Guo–Krasnosel'skiĭ fixed point theorem.

Lemma 3.1. *Suppose* (H0) *and* (H1) *hold and let* $\rho^* > 0$. *Then there is a* $\Lambda > 0$ *such that, for every* $\lambda \geq \Lambda$ *and* $(a, b) \in [0, \infty)^2$,

$$\|T(\bar{u}, \bar{v})\| \geq \|(\bar{u}, \bar{v})\|$$

for $(\bar{u}, \bar{v}) \in C \cap \partial \Omega_{\rho^*}$.

Proof. Let $\rho^* > 0$ and let $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho^*}$. Let $r = \alpha(1-\beta)$, where α and β are as in (H1) and note $r \in (0, 1)$. Furthermore, choose $c \geq 1$ so that both $\bar{u}' + a \leq c\|\bar{u}'\|_\infty$ and $\bar{v}' + b \leq c\|\bar{v}'\|_\infty$ hold for $t \in [\alpha, \beta]$. Define

$$M = \inf\left\{ \frac{f(t, ra_1, ca_2, ra_3, ca_4)}{r(a_1 + a_3) + c(a_2 + a_4)} : t \in [\alpha, \beta], \ a_1, a_2, a_3 > 0, \ a_4 \geq 0, \right.$$
$$\left. \text{and} \ a_1 + a_2 + a_3 + a_4 = p^* \right\}.$$

The existence of a positive M follows from (H1). Set $\Lambda \geq \left[Mr \int_\alpha^\beta G(1, s)\, ds \right]^{-1}$.

As $(\bar{u}, \bar{v}) \in C$, by Lemma 2.2, we have $\bar{u}(t) + at \geq \bar{u}(t) \geq r\|\bar{u}\|_\infty$. Moreover, due to the nondecreasing property of f in the second and fourth variables and its nonincreasing property in the third and fifth variables, we see that

$$\|T(\bar{u}, \bar{v})\| \geq \|A_2(\bar{u}, \bar{v})\|_\infty$$

$$\geq \lambda \int_0^1 G(1, s) f\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds$$

$$\geq \lambda \int_\alpha^\beta G(1, s) f\left(s, r\|\bar{u}\|_\infty, c\|\bar{u}'\|_\infty, r\|\bar{v}\|_\infty, c\|\bar{v}'\|_\infty\right) ds$$

$$\geq \lambda\left[r(\|\bar{u}\|_\infty + \|\bar{v}\|_\infty) + c(\|\bar{u}'\|_\infty + \|\bar{v}'\|_\infty)\right]$$
$$\times \int_\alpha^\beta G(1, s) \frac{f\left(s, r\|\bar{u}\|_\infty, c\|\bar{u}'\|_\infty, r\|\bar{v}\|_\infty, c\|\bar{v}'\|_\infty\right)}{r(\|\bar{u}\|_\infty + \|\bar{v}\|_\infty) + c(\|\bar{u}'\|_\infty + \|\bar{v}'\|_\infty)} ds$$

$$\geq \lambda M\left[r(\|\bar{u}\|_\infty + \|\bar{v}\|_\infty) + c(\|\bar{u}'\|_\infty + \|\bar{v}'\|_\infty)\right] \int_\alpha^\beta G(1, s)\, ds$$

$$\geq \lambda Mr\left(\|\bar{u}\|_\infty + \|\bar{u}'\|_\infty + \|\bar{v}\|_\infty + \|\bar{v}'\|_\infty\right) \int_\alpha^\beta G(1, s)\, ds$$

$$\geq \lambda Mr\|(\bar{u}, \bar{v})\| \int_\alpha^\beta G(1, s)\, ds$$

$$\geq \Lambda Mr\|(\bar{u}, \bar{v})\| \int_\alpha^\beta G(1, s)\, ds$$

$$\geq \|(\bar{u}, \bar{v})\|. \qquad \square$$

Lemma 3.2. *Fix $\Lambda > 0$. Suppose* (H0) *and* (H1) *hold. Then, for all $\lambda \geq \Lambda$ and for all $(a, b) \in [0, \infty)^2$, with $a + b > 0$, there exists a $\rho_1 = \rho_1(\Lambda, a, b)$ such that for every $\rho \in (0, \rho_1)$, we have*

$$\|T(\bar{u}, \bar{v})\| \geq \|(\bar{u}, \bar{v})\|$$

for all $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_\rho$.

Proof. Fix $\Lambda > 0$. By (H1) and the nonincreasing/nondecreasing properties of f, there exists $k > 0$ such that

$$f\left(t, \bar{u} + ta, \bar{u}' + a, \bar{v} + tb, \bar{v}' + b\right) \geq f\left(t, \alpha a, \|\bar{u}'\|_\infty + a, \alpha b, \|\bar{v}'\|_\infty + b\right) > k$$

for all $t \in (\alpha, \beta)$, where α and β are as in (H1). Take $\rho_1 = \Lambda k \int_\alpha^\beta G(1, s)\, ds$. Then, for $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_\rho$ where $\rho \leq \rho_1$,

$$
\begin{aligned}
\|T(\bar{u}, \bar{v})\| \geq \|A_2(\bar{u}, \bar{v})\|_\infty &\geq \lambda \int_0^1 G(1, s) f\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds \\
&\geq \lambda \int_\alpha^\beta G(1, s) f\left(s, \alpha a \|\bar{u}'\|_\infty + a, \alpha b, \|\bar{v}'\|_\infty + b\right) ds \\
&> \lambda k \int_\alpha^\beta G(1, s)\, ds \\
&= \lambda k \|(\bar{u}, \bar{v})\| \int_\alpha^\beta \frac{G(1, s)}{\|(\bar{u}, \bar{v})\|}\, ds \\
&\geq \Lambda k \|(\bar{u}, \bar{v})\| \int_\alpha^\beta \frac{G(1, s)}{\|(\bar{u}, \bar{v})\|}\, ds \\
&= \frac{\rho_1}{\rho} \|(\bar{u}, \bar{v})\| \\
&\geq \|(\bar{u}, \bar{v})\|. \qquad \square
\end{aligned}
$$

Lemma 3.3. *Suppose* (H0), (H2) *and* (H4) *hold and let $\rho^* > 0$ be fixed. Then given $\lambda > 0$, there is a $\rho_2 \in (0, \rho^*)$ and a $\delta > 0$ such that for every $(a, b) \in [0, \infty)^2$, with $0 < a + b < \delta$, we have*

$$\|T(\bar{u}, \bar{v})\| \leq \|(\bar{u}, \bar{v})\|$$

for $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_2}$.

Proof. Let $\lambda > 0$. Pick $\epsilon > 0$ so that $\lambda\epsilon < \frac{1}{3}$. Then, by (H2), we can find a $\rho_2 \in (0, \rho^*)$ such that, for all $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 = \rho_2$ and $a + b \leq \rho_2$ with $\rho_2 < \frac{1}{2}q$, where $q > 0$ is as in (H4), we have

$$f(t, x_1 + a, x_2, x_3 + b, x_4) < \epsilon\left[(x_1 + a) + x_2 + (x_3 + b) + x_4\right]$$

for $t \in [0, 1]$.

Take $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_2}$, and suppose $a + b \leq \rho_2$. Notice that there exists $c \in (0, 1]$ such that $\bar{u}' + a \geq c\|\bar{u}'\|_\infty$ and $\bar{v}' + b \geq c\|\bar{v}'\|_\infty$. Then, for $t \in [0, 1]$, we have

$$A_2(\bar{u}, \bar{v})(t) = \lambda \int_0^1 G(t, s) f\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds$$

$$\leq \lambda \int_0^1 G(t, s) f\left(s, \|\bar{u}\|_\infty + a, c\|\bar{u}'\|_\infty, \|\bar{v}\|_\infty + b, c\|\bar{v}'\|_\infty\right) ds$$

$$< \lambda\epsilon\left[\|\bar{u}\|_\infty + c\|\bar{u}'\|_\infty + \|\bar{v}\|_\infty + c\|\bar{v}'\|_\infty + (a+b)\right] \int_0^1 G(t, s)\, ds$$

$$\leq \lambda\epsilon\left[\|(\bar{u}, \bar{v})\| + (a+b)\right] \int_0^1 G(t, s)\, ds$$

$$\leq 2\lambda\epsilon\|(\bar{u}, \bar{v})\| \int_0^1 G(t, s)\, ds$$

$$\leq \lambda\epsilon\|(\bar{u}, \bar{v})\|.$$

Using a similar argument to the one above, we see that

$$A_2'(\bar{u}, \bar{v})(t) = \lambda \int_0^1 \frac{\partial}{\partial t} G(t, s) f\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds$$

$$\leq 2\lambda\epsilon\|(\bar{u}, \bar{v})\| \int_0^1 \frac{\partial}{\partial t} G(t, s)\, ds$$

$$\leq 2\lambda\epsilon\|(\bar{u}, \bar{v})\|.$$

In other words,

$$\|A_2(\bar{u}, \bar{v})\|_\infty + \|A_2'(\bar{u}, \bar{v})\|_\infty \leq 3\lambda\epsilon\|(\bar{u}, \bar{v})\|.$$

By (H4), since $\left[(\|\bar{u}\|_\infty + a) + \|\bar{u}'\|_\infty + (\|\bar{v}\|_\infty + b) + \|\bar{v}'\|_\infty\right] \leq 2\rho_2 < q$, we have

$$g\left(t, \|\bar{u}\|_\infty + a, \|\bar{u}'\|_\infty, \|\bar{v}\|_\infty + b, \|\bar{v}'\|_\infty\right)$$

$$\leq \gamma\left(\|\bar{u}\|_\infty + a + \|\bar{u}'\|_\infty + \|\bar{v}\|_\infty + b + \|\bar{v}'\|_\infty\right).$$

Let $\delta' < 1$ and set $\delta = \delta'\rho_2$. Then for $a + b < \delta$, $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_2}$, and $t \in [0, 1]$, we have

$$A_1(\bar{u}, \bar{v})(t) = \int_0^1 G(t, s) g\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds$$

$$\leq \int_0^1 G(t, s) g\left(s, \|\bar{u}\|_\infty + a, c\|\bar{u}'\|_\infty, \|\bar{v}\|_\infty + b, c\|\bar{v}'\|_\infty\right) ds$$

$$\leq \gamma\left[\|\bar{u}\|_\infty + c\|\bar{u}'\|_\infty + \|\bar{v}\|_\infty + c\|\bar{v}'\|_\infty + (a+b)\right] \int_0^1 G(t, s)\, ds$$

$$\leq \gamma\left[\|(\bar{u}, \bar{v})\| + (a+b)\right] \int_0^1 G(t, s)\, ds$$

$$< \gamma(1+\delta')\|(\bar{u}, \bar{v})\| \int_0^1 G(t, s)\, ds$$

$$\leq \tfrac{1}{2}\gamma(1+\delta')\|(\bar{u}, \bar{v})\|,$$

where c is as above. And similarly,

$$A_1'(\bar{u}, \bar{v})(t) = \int_0^1 \frac{\partial}{\partial t} G(t, s) g\big(s, \bar{u}+sa, \bar{u}'+a, \bar{v}+sb, \bar{v}'+b\big)\, ds$$

$$< \gamma(1+\delta')\|(\bar{u}, \bar{v})\| \int_0^1 \frac{\partial}{\partial t} G(t, s)\, ds$$

$$\leq \gamma(1+\delta')\|(\bar{u}, \bar{v})\|.$$

Hence,

$$\|A_1(\bar{u}, \bar{v})\|_\infty + \|A_1'(\bar{u}, \bar{v})\|_\infty < \tfrac{3}{2}\gamma(1+\delta')\|(\bar{u}, \bar{v})\|.$$

Thus, for $a + b < \delta$, we have

$$\|T(\bar{u}, \bar{v})\| = \|A_1(\bar{u}, \bar{v})\|_\infty + \|A_1'(\bar{u}, \bar{v})\|_\infty + \|A_2(\bar{u}, \bar{v})\|_\infty + \|A_2'(\bar{u}, \bar{v})\|_\infty$$

$$< \big[\tfrac{3}{2}\gamma(1+\delta') + 3\lambda\epsilon\big]\|(\bar{u}, \bar{v})\|.$$

For small enough ϵ and δ', it follows that $\|T(\bar{u}, \bar{v})\| \leq \|(\bar{u}, \bar{v})\|$.   □

Lemma 3.4. *Let $\delta > 0$. Suppose $0 < a + b < \delta$ and (H0), (H3) and (H5) hold. Then, for every $\lambda > 0$, there is a $\rho_3 = \rho_3(\delta, \lambda)$ such that for all $\rho \geq \rho_3$,*

$$\|T(\bar{u}, \bar{v})\| \leq \|(\bar{u}, \bar{v})\|,$$

where $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_\rho$.

Proof. Let $\delta > 0$, $0 < a + b < \delta$ and let $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$. By (H5) and the nondecreasing/nonincreasing properties of g as in (H0), given any $q_1 \geq \hat{\rho}$, we have

$$g(t, x_1+a, x_2, x_3+a, x_4) \leq \eta(x_1+a+x_2+x_3+b+x_4)$$

for $x_1 + x_2 + x_3 + x_4 \geq q_1$ and $t \in [0, 1]$.

Let $\epsilon > 0$ and pick $q_1 \geq \hat{\rho}$ large enough so that $\epsilon > \eta\delta/q_1$. Let $x_1+x_2+x_3+x_4 \geq q_1$. Then

$$g(t, x_1+a, x_2, x_3+a, x_4) \leq \eta(x_1+x_2+x_3+x_4) + \eta(a+b)$$

$$< \eta(x_1+x_2+x_3+x_4) + \epsilon(x_1+x_2+x_3+x_4)$$

$$= (\eta+\epsilon)(x_1+x_2+x_3+x_4).$$

Let $(\bar{u}, \bar{v}) \in C \cap \partial \Omega_{q_1}$. Pick $c \in (0, 1]$ such that $\bar{u}' + a \geq c \|\bar{u}'\|_\infty$ and $\bar{v}' + b \geq c \|\bar{v}'\|_\infty$. Then for $t \in [0, 1]$,

$$A_1(\bar{u}, \bar{v})(t) = \int_0^1 G(t, s) g\big(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\big) ds$$

$$\leq \int_0^1 G(t, s) g\big(s, \|\bar{u}\|_\infty + a, c\|\bar{u}'\|_\infty, \|\bar{v}\|_\infty + b, c\|\bar{v}'\|_\infty\big) ds$$

$$< (\eta + \epsilon) \|(\bar{u}, \bar{v})\| \int_0^1 G(t, s) ds.$$

A similar argument shows that

$$A_1'(\bar{u}, \bar{v})(t) = \int_0^1 \frac{\partial}{\partial t} G(t, s) g\big(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\big) ds$$

$$< (\eta + \epsilon) \|(\bar{u}, \bar{v})\| \int_0^1 \frac{\partial}{\partial t} G(t, s) ds.$$

Combining these inequalities, we see that

$$\|A_1(\bar{u}, \bar{v})\|_\infty + \|A_1'(\bar{u}, \bar{v})\|_\infty < \tfrac{3}{2}(\eta + \epsilon) \|(\bar{u}, \bar{v})\|.$$

Now consider $A_2(\bar{u}, \bar{v})(t)$. Let $\delta' > 0$. Then, by (H0) and (H3), there is a $q_2 > 0$ such that for all $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 \geq q_2$, we have

$$f(t, x_1 + a, x_2, x_3 + b, x_4) \leq \delta'(x_1 + a + x_2 + x_3 + b + x_4)$$

for every $t \in [0, 1]$. Let $q_3 = \max\{\delta, q_2\}$. Noting that $a + b < \delta$, for $(x_1, x_2, x_3, x_4) \in [0, \infty)^4$ with $x_1 + x_2 + x_3 + x_4 \geq q_3$, we have

$$f(t, x_1 + a, x_2, x_3 + b, x_4) \leq \delta' \big[(x_1 + x_2 + x_3 + x_4) + q_3\big]$$

$$\leq 2\delta'(x_1 + x_2 + x_3 + x_4).$$

Then for $t \in [0, 1]$ and any $(\bar{u}, \bar{v}) \in C \cap \partial \Omega_{q_3}$,

$$A_2(\bar{u}, \bar{v}) = \lambda \int_0^1 G(t, s) f\big(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\big) ds$$

$$\leq \lambda \int_0^1 G(t, s) f\big(s, \|\bar{u}\|_\infty + a, c\|\bar{u}'\|_\infty, \|\bar{v}\|_\infty + b, c\|\bar{v}'\|_\infty\big) ds$$

$$< \lambda \cdot 2\delta' \|(\bar{u}, \bar{v})\| \int_0^1 G(t, s) ds,$$

where c is as above. And similarly,

$$A_2(\bar{u}, \bar{v}) = \lambda \int_0^1 \frac{\partial}{\partial t} G(t, s) f\left(s, \bar{u} + sa, \bar{u}' + a, \bar{v} + sb, \bar{v}' + b\right) ds$$

$$< \lambda \cdot 2\delta' \|(\bar{u}, \bar{v})\| \int_0^1 \frac{\partial}{\partial t} G(t, s) ds.$$

Combining these inequalities, we see that

$$\|A_2(\bar{u}, \bar{v})\|_\infty + \|A_2'(\bar{u}, \bar{v})\|_\infty < 3\lambda\delta' \|(\bar{u}, \bar{v})\|.$$

Take $\rho_3 = \max\{q_1, q_3\}$ and let $\rho \geq \rho_3$. Then given $(\bar{u}, \bar{v}) \in C \cap \partial\Omega_\rho$, we see that

$$\|T(\bar{u}, \bar{v})\| = \|A_1(\bar{u}, \bar{v})\|_\infty + \|A_1'(\bar{u}, \bar{v})\|_\infty + \|A_2(\bar{u}, \bar{v})\|_\infty + \|A_2'(\bar{u}, \bar{v})\|_\infty$$

$$< \left[\tfrac{1}{2}(6\lambda\delta' + 3(\eta + \epsilon))\right] \|(\bar{u}, \bar{v})\|.$$

Recall by (H5) that $\eta \in \left(0, \frac{2}{3}\right)$. Pick ϵ and δ' small enough that $6\lambda\delta' + 3\epsilon \leq 2 - 3\eta$. Thus, we have the desired result. $\qquad\square$

4. The main result

Theorem 4.1. *Let continuous functions* $f, g : [0, 1] \times [0, \infty)^4 \to [0, \infty)$ *satisfy hypotheses* (H0)–(H5). *Then there exists* $\Lambda > 0$ *such that given* $\lambda \geq \Lambda$, *there exists* $\delta > 0$ *such that for every* $a, b \geq 0$ *satisfying* $0 < a + b < \delta$, *the system* (5)–(8) *has at least three positive solutions.*

Proof. Suppose f, g satisfy hypotheses (H0)–(H5). Let $\rho^* > 0$ be fixed. By Lemma 3.1, there is $\Lambda > 0$ such that, for every $\lambda \geq \Lambda$ and $a, b \geq 0$,

$$\|T(\bar{u}, \bar{v})\| \geq \|(\bar{u}, \bar{v})\| \quad \text{for } (\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho^*}.$$

Now, fix $\lambda \geq \Lambda$. Lemmas 3.2 through 3.4 give that there is $\delta > 0$ and $\rho_1, \rho_2, \rho_3 > 0$ satisfying $\rho_1 < \rho_2 < \rho^* < \rho_3$ such that for $(a, b) \in [0, \infty)^2$ with $0 < a + b < \delta$,

$$\|T(\bar{u}, \bar{v})\| \geq \|(\bar{u}, \bar{v})\| \quad \text{for } (\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_1},$$

$$\|T(\bar{u}, \bar{v})\| \leq \|(\bar{u}, \bar{v})\| \quad \text{for } (\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_2},$$

$$\|T(\bar{u}, \bar{v})\| \leq \|(\bar{u}, \bar{v})\| \quad \text{for } (\bar{u}, \bar{v}) \in C \cap \partial\Omega_{\rho_3}.$$

Applying the Guo–Krasnosel'skiĭ fixed point theorem three times, we get the existence of three positive solutions, $(\bar{u}_1, \bar{v}_1), (\bar{u}_2, \bar{v}_2), (\bar{u}_3, \bar{v}_3) \in C$ such that

$$\rho_1 < \|(\bar{u}_1, \bar{v}_1)\| < \rho_2 < \|(\bar{u}_2, \bar{v}_2)\| < \rho^* < \|(\bar{u}_3, \bar{v}_3)\| < \rho_3. \qquad\square$$

Recall that solutions to the system (5)–(8) are in one-to-one correspondence with those of the system (1)–(4). Thus we have our desired result.

References

[Agarwal 1989] R. P. Agarwal, "On fourth order boundary value problems arising in beam analysis", *Differential Integral Equations* **2**:1 (1989), 91–110. MR Zbl

[Henderson and Hopkins 2010] J. Henderson and B. Hopkins, "Multiple positive solutions for a discrete fourth order nonhomogeneous boundary value problem", *Int. Electron. J. Pure Appl. Math.* **2**:2 (2010), 81–92.

[Hopkins 2009] B. Hopkins, *Multiplicity of positive solutions of even-order nonhomogeneous boundary value problems*, thesis, Baylor University, Waco, TX, 2009, http://search.proquest.com/docview/304849906. MR

[Hopkins 2015] B. Hopkins, "Multiplicity of positive solutions for an even-order right focal boundary value problem", *Adv. Dyn. Syst. Appl.* **10**:2 (2015), 189–200. MR

[Krasnosel'skiĭ 1964] M. A. Krasnosel'skiĭ, *Positive solutions of operator equations*, P. Noordhoff Ltd., Groningen, 1964. MR Zbl

[do Ó et al. 2008] J. M. do Ó, S. Lorca, and P. Ubilla, "Multiplicity of solutions for a class of non-homogeneous fourth-order boundary value problems", *Appl. Math. Lett.* **21**:3 (2008), 279–286. MR Zbl

oliviabennett15@gmail.com   *Department of Mathematics and Statistics, University of Central Oklahoma, 100 N. University Drive, Edmond, OK 73034, United States*

dbrumley1@uco.edu   *Department of Mathematics and Statistics, University of Central Oklahoma, 100 N. University Drive, Edmond, OK 73034, United States*

bhopkins3@uco.edu   *Department of Mathematics and Statistics, University of Central Oklahoma, 100 N. University Drive, Edmond, OK 73034, United States*

kkarber1@uco.edu   *Department of Mathematics and Statistics, University of Central Oklahoma, 100 N. University Drive, Edmond, OK 73034, United States*

tmilligan1@uco.edu   *Department of Mathematics and Statistics, University of Central Oklahoma, 100 N. University Drive, Edmond, OK 73034, United States*

Factorization of Temperley–Lieb diagrams

Dana C. Ernst, Michael G. Hastings and Sarah K. Salmon

(Communicated by Scott T. Chapman)

The Temperley–Lieb algebra is a finite-dimensional associative algebra that arose in the context of statistical mechanics and occurs naturally as a quotient of the Hecke algebra arising from a Coxeter group of type *A*. It is often realized in terms of a certain diagram algebra, where every diagram can be written as a product of "simple diagrams". These factorizations correspond precisely to factorizations of the so-called fully commutative elements of the Coxeter group that index a particular basis. Given a reduced factorization of a fully commutative element, it is straightforward to construct the corresponding diagram. On the other hand, it is generally difficult to reconstruct the factorization given an arbitrary diagram. We present an efficient algorithm for obtaining a reduced factorization for a given diagram.

1. Introduction

The Temperley–Lieb algebra [1971] is a finite-dimensional associative algebra that arose in the context of statistical mechanics. Penrose [1971] and Kauffman [1990] showed that this algebra can be faithfully represented by a diagram algebra that has a basis given by certain diagrams. Jones [1999] showed that the Temperley–Lieb algebra occurs naturally as a quotient of the Hecke algebra arising from a Coxeter group of type *A* (whose underlying group is the symmetric group). This realization of the Temperley–Lieb algebra as a Hecke algebra quotient was later generalized to the case of an arbitrary Coxeter group by Graham [1995]. These generalized Temperley–Lieb algebras have a basis indexed by the fully commutative elements (in the sense of Stembridge [1996]) of the underlying Coxeter group. In cases when diagrammatic representations are known to exist, it turns out that every diagram can be written as a product of "simple diagrams". Each factorization of a diagram corresponds precisely to a factorization of the fully commutative element that indexes the diagram. Given a diagrammatic representation and a reduced factorization of a fully commutative element, it is easy to construct the corresponding diagram. However, given an arbitrary basis diagram, it is generally

difficult to reconstruct the factorization of the corresponding group element. In the (type A) Temperley–Lieb algebra, we have devised an algorithm for obtaining a reduced factorization for a given diagram.

This paper is organized as follows. In Section 2, we recall the basic terminology of Coxeter groups, fully commutative elements, heaps, and the Temperley–Lieb algebra, as well as establish our notation and review several necessary results. Section 3 describes the construction of the diagram algebra that is a faithful representation of the Temperley–Lieb algebra. This section includes a description of both the so-called simple diagrams that generate the algebra, as well as the basis that is indexed by the fully commutative elements of the Coxeter group of type A. We present our algorithm for factoring a given Temperley–Lieb diagram in terms of the heap associated to the corresponding fully commutative element in Section 4. We conclude with Section 5, which details potential further research.

2. Preliminaries

Coxeter groups. A *Coxeter system* is a pair (W, S) consisting of a finite set S of generating involutions and a group W, called a *Coxeter group*, with presentation

$$W = \langle S \mid (st)^{m(s,t)} = e \text{ for } m(s, t) < \infty \rangle,$$

where e is the identity, $m(s, t) = 1$ if and only if $s = t$, and $m(s, t) = m(t, s)$. It follows that the elements of S are distinct as group elements and that $m(s, t)$ is the order of st [Humphreys 1990]. Coxeter groups are generalizations of reflection groups, where each generator $s \in S$ can be thought of as a reflection. Recall that the composition of two reflections is a rotation by twice the angle between the corresponding hyperplanes. So if $s, t \in S$, we can think of st as a rotation with order $m(s, t)$.

Since elements of S have order 2, the relation $(st)^{m(s,t)} = e$ can be written as

$$\underbrace{sts\cdots}_{m(s,t)} = \underbrace{tst\cdots}_{m(s,t)} \tag{1}$$

with $m(s, t) \geq 2$ factors. If $m(s, t) = 2$, then $st = ts$ is called a *commutation relation* since s and t commute. Otherwise, if $m(s, t) \geq 3$, then the corresponding relation is called a *braid relation*. The replacement

$$\underbrace{sts\cdots}_{m(s,t)} \mapsto \underbrace{tst\cdots}_{m(s,t)}$$

will be referred to as a *commutation* if $m(s, t) = 2$ and a *braid move* if $m(s, t) \geq 3$.

We can represent the Coxeter system (W, S) with a unique *Coxeter graph* Γ with

(1) vertex set $S = \{s_1, \ldots, s_n\}$ and

(2) edges $\{s_i, s_j\}$ for each $m(s_i, s_j) \geq 3$.
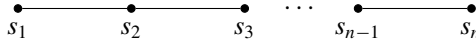
Figure 1. Coxeter graph of type A_n.

Each edge $\{s_i, s_j\}$ is labeled with its corresponding bond strength $m(s_i, s_j)$. Since bond strength 3 is the most common, we typically omit the labels of 3 on those edges.

There is a one-to-one correspondence between Coxeter systems and Coxeter graphs. Given a Coxeter graph Γ, we can construct the corresponding Coxeter system (W, S). In this case, we say that (W, S), or just W, is of type Γ. If (W, S) is of type Γ, for emphasis, we may write (W, S) as $(W(\Gamma), S(\Gamma))$. Note that generators s_i and s_j are connected by an edge in the Coxeter graph Γ if and only if s_i and s_j do not commute [Humphreys 1990].

The Coxeter system of type A_n is given by the Coxeter graph in Figure 1. In this case, $W(A_n)$ is generated by $S(A_n) = \{s_1, s_2, \ldots, s_n\}$ and has defining relations

(1) $s_i s_i = e$ for all i;

(2) $s_i s_j = s_j s_i$ when $|i - j| > 1$;

(3) $s_i s_j s_i = s_j s_i s_j$ when $|i - j| = 1$.

The Coxeter group $W(A_n)$ is isomorphic to the symmetric group S_{n+1} under the mapping that sends s_i to the adjacent transposition $(i, i+1)$. This paper focuses on an associative algebra whose underlying structure is a Coxeter system of type A_n.

Let S^* denote the free monoid over S. If a word $\mathsf{w} = s_{x_1} s_{x_2} \cdots s_{x_m} \in S^*$ is equal to w when considered as an element of W, we say that w is an *expression* for w. (Expressions will be written in sans serif font for clarity.) Furthermore, if m is minimal among all possible expressions for w, we say that w is a *reduced expression* for w, and we call m the *length* of w, denoted $\ell(w)$. Each element $w \in W$ can have several different reduced expressions that represent it. The following theorem, called Matsumoto's theorem [Geck and Pfeiffer 2000], indicates how all of the reduced expressions for a given group element are related.

Proposition 2.1. *In a Coxeter group W, any two reduced expressions for the same group element differ by a finite sequence of commutations and braid moves.* $\square$

Let w be a reduced expression for $w \in W$. We define a *subexpression* of w to be any subsequence of w. We will refer to a consecutive subexpression of w as a *subword*.

Example 2.2. Let $\mathsf{w} = s_1 s_2 s_4 s_5 s_2 s_6 s_5$ be an expression for $w \in W(A_6)$. Then

$$s_1 {\color{blue}s_2 s_4} s_5 s_2 s_6 s_5 = s_1 s_4 {\color{blue}s_2 s_5} s_2 s_6 s_5 = s_1 s_4 s_5 {\color{green}s_2 s_2} s_6 s_5 = s_1 s_4 s_5 s_6 s_5,$$

where the blue subword indicates the location where a commutation is applied to obtain the next expression and the green subword indicates the location where

two adjacent occurrences of the same generator are canceled to obtain the last expression. This shows that w is not reduced. It turns out that $s_1 s_4 s_5 s_6 s_5$ is a reduced expression for w and hence $\ell(w) = 5$.

Example 2.3. Let w $= s_1 s_2 s_3 s_4 s_2$ be a reduced expression for $w \in W(A_4)$. Then the set of all reduced expressions for w is given by

$$\{s_1 s_2 s_3 s_4 s_2, \; s_1 s_2 s_3 s_2 s_4, \; s_1 s_3 s_2 s_3 s_4, \; s_3 s_1 s_2 s_3 s_4\},$$

where the blue subwords indicate the location where a commutation is applied to obtain the next expression in the set and the pink subword indicates the location where a braid move is applied to obtain the third expression from the second expression. Note that $\ell(w) = 5$.

Fully commutative elements. Let (W, S) be a Coxeter system of type Γ and let $w \in W$. Following [Stembridge 1996], we define a relation $\sim$ on the set of reduced expressions for w. Let w and w$'$ be two reduced expressions for w. We define w $\sim$ w$'$ if we can obtain w$'$ from w by applying a single commutation move of the form $st \mapsto ts$, where $m(s, t) = 2$. Now, define the equivalence relation $\approx$ by taking the reflexive transitive closure of $\sim$. Each equivalence class under $\approx$ is called a *commutation class*. Two reduced expressions are said to be *commutation equivalent* if they are in the same commutation class.

Example 2.4. Let w $= s_1 s_2 s_3 s_4 s_5 s_2$ and w$' = s_1 s_2 s_3 s_2 s_4 s_5$ be two different reduced expressions for $w \in W(A_5)$. Then w and w$'$ are commutation equivalent since

$$s_1 s_2 s_3 s_4 s_5 s_2 = s_1 s_2 s_3 s_4 s_2 s_5 = s_1 s_2 s_3 s_2 s_4 s_5,$$

where the blue subwords indicate the location where a commutation is applied to obtain the next expression. By applying a braid relation to w$'$, we obtain

$$s_1 s_2 s_3 s_2 s_4 s_5 = s_1 s_3 s_2 s_3 s_4 s_5,$$

where the location of the braid move has been highlighted in pink. It turns out that the last reduced expression above is neither commutation equivalent to w nor w$'$, and hence w has more than one commutation class. Specifically, the commutation classes are

$$\{s_1 s_2 s_3 s_4 s_5 s_2, \; s_1 s_2 s_3 s_4 s_2 s_5, \; s_1 s_2 s_3 s_2 s_4 s_5\} \quad \text{and} \quad \{s_1 s_3 s_2 s_3 s_4 s_5, \; s_3 s_1 s_2 s_3 s_4 s_5\}.$$

Example 2.5. Let w $= s_2 s_1 s_3 s_4 s_2$ be a reduced expression for $w \in W(A_4)$. In this case, w has exactly five reduced expressions, including w. From this, it is easy to verify that all reduced expressions for w are commutation equivalent. This implies that there is a unique commutation class for w:

$$\{s_2 s_1 s_3 s_4 s_2, \; s_2 s_3 s_1 s_4 s_2, \; s_2 s_1 s_3 s_2 s_4, \; s_2 s_3 s_1 s_2 s_4, \; s_2 s_3 s_4 s_1 s_2\}.$$

If w has exactly one commutation class, then we say that w is *fully commutative*, or just FC. The set of all fully commutative elements of W is denoted by $\mathrm{FC}(\Gamma)$, where Γ is the corresponding Coxeter graph. For consistency, we say that a reduced expression w is FC if it is a reduced expression for some $w \in \mathrm{FC}(\Gamma)$. Note that the element in Example 2.4 is not FC since there are two commutation classes, while the element in Example 2.5 is FC.

Given some $w \in \mathrm{FC}(\Gamma)$ and a starting reduced expression for w, observe that the definition of fully commutative states that one only needs to perform commutations to obtain all the reduced expression for w, but the following result due to Stembridge [1996] states that when w is FC, performing commutations is the only possible way to obtain another reduced expression for w.

Proposition 2.6. *An element $w \in W$ is FC if and only if no reduced expression for w contains*

$$\underbrace{sts \cdots}_{m(s,t)}$$

as a subword when $m(s, t) \geq 3$. □

In other words, an element is FC if and only if there is no opportunity to apply a braid move. For example, we can conclude that the element in Example 2.4 is not FC without actually computing the commutation classes since there is an opportunity to apply a braid move, which we highlighted in pink.

Stembridge classified the irreducible Coxeter groups that contain only finitely many fully commutative elements, called the *FC-finite Coxeter groups*. This paper is mainly concerned with $W(A_n)$, which is a finite group, so it has finitely many FC elements. However, there exist infinite Coxeter groups that contain only finitely many FC elements. For example, Coxeter groups of type E_n with $n \geq 9$ are infinite, but they have only finitely many FC elements. It is well known that the number of FC elements in $W(A_n)$ is equal to the $(n+1)$-th Catalan number, where the k-th Catalan number is given by

$$C_k = \frac{1}{k+1} \binom{2k}{k}.$$

Heaps. Each reduced expression is associated with a labeled partially ordered set called a heap. Heaps provide a visual representation of the reduced expression while preserving the relations of the generators. We follow the development in [Ernst 2010; Stembridge 1996].

Let (W, S) be a Coxeter system. Suppose $\mathsf{w} = s_{x_1} s_{x_2} \cdots s_{x_k}$ is a reduced expression for $w \in W$, and as in [Stembridge 1996], define a partial ordering $\prec$ on the indices $\{1, \ldots, k\}$ by the transitive closure of the relation $j \prec i$ if $i < j$ and s_{x_i} and s_{x_j} do not commute. In particular, $j \prec i$ if $i < j$ and $s_{x_i} = s_{x_j}$, by transitivity and the fact that w is reduced. This partial order with i labeled s_{x_i} is called the *heap* of w. Note

Figure 2. Labeled Hasse diagram and lattice point representation of a heap.

that for simplicity, we are omitting the labels of the underlying poset but retaining the labels of the corresponding generators.

Example 2.7. Let $w = s_2 s_1 s_3 s_2 s_4 s_5$ be a reduced expression for $w \in W(A_5)$. Since $\ell(w) = 6$, the expression w is indexed by $\{1, 2, 3, 4, 5, 6\}$. We see that $4 \prec 3$ since $3 < 4$ and the third and fourth factors (namely, s_3 and s_2) do not commute. The labeled Hasse diagram for the heap of w is shown in Figure 2 (left).

Let w be a fixed reduced expression for $w \in W(A_n)$. As in [Billey and Jones 2007; Ernst 2010], we represent a heap for w as a set of lattice points embedded in $\{1, \ldots, n\} \times \mathbb{N}$. To do so, we assign coordinates $(x, y) \in \{1, \ldots, n\} \times \mathbb{N}$ to each entry of the labeled Hasse diagram for the heap of w in such a way that

(1) an entry labeled s_i in the heap has coordinates (x, y) if and only if $x = i$;

(2) an entry with coordinates (x, y) is greater than an entry with coordinates (x', y') in the heap if and only if $y > y'$.

It follows from the definition that there is an edge in the Hasse diagram from (x, y) to (x', y') if and only if $x = x' \pm 1$, $y > y'$, and there are no entries (x'', y'') such that $x'' \in \{x, x'\}$ and $y' < y'' < y$. This implies that we can completely reconstruct the edges of the Hasse diagram and the corresponding heap poset from a lattice point representation.

Note that our heaps are upside-down versions of the heaps that appear in [Billey and Jones 2007] and several other papers. That is, in this paper, entries on top of a heap correspond to generators occurring to the left, as opposed to the right, in the corresponding reduced expression. One can form similar lattice point representations for heaps when Γ is a straight line Coxeter graph.

Let $w = s_{x_1} \cdots s_{x_k}$ be any reduced expression for $w \in W(A_n)$. We let $H(w)$ denote a lattice representation of the heap poset in $\{1, \ldots, n\} \times \mathbb{N}$ described in the paragraph above. There are many possible coordinate assignments for the entries of $H(w)$, yet the x-coordinates for each entry will be fixed. If s_{x_i} and s_{x_j} are adjacent generators in the Coxeter graph with $i < j$, then we must place the point labeled by s_{x_i} at a level that is *above* the level of the point labeled by s_{x_j}. In particular, two entries labeled

by the same generator may only differ by the amount of vertical space between them while maintaining their relative vertical position to adjacent entries in the heap.

Because generators that are not adjacent in the Coxeter graph commute, points whose x-coordinates differ by more than 1 can slide past each other or land at the same level. To visualize the labeled heap poset of a lattice representation we will enclose each entry of the heap in a block in such a way that if one entry covers another, the blocks overlap halfway.

It follows from [Stembridge 1996, Proposition 2.2] that heaps are well-defined up to commutation class. In particular, there is a one-to-one correspondence between commutation classes and heaps. That is, if w and w′ are two reduced expressions for $w \in W$ that are in the same commutation class, then the heaps of w and w′ are equal. Conversely, if w and w′ belong to different commutation classes, then the corresponding heaps will be different. In particular, if w is FC, then it has a single commutation class, and so there is a unique heap associated to w. In this case, if w is FC, then we may write $H(w)$ to denote the heap of any reduced expression for w. We will not make a distinction between $H(w)$ and its lattice point representation.

There are potentially many different ways to represent a heap, each differing by the vertical placement of blocks. For example, we can place blocks in vertical positions that are as high as possible, as low as possible, or some combination of high and low. When w is FC, we wish to make a canonical choice for the representation of $H(w)$ by giving all blocks at the top of the heap the same vertical position and placing all other blocks as high as possible. Note that our canonical representation of heaps of FC elements corresponds precisely to the unique heap factorization of [Viennot 1986, Lemma 2.9] and to the Cartier–Foata normal form for monomials [Cartier and Foata 1969; Green 2006].

Example 2.8. The canonical lattice point representation of $H(w)$ for the reduced expression given in Example 2.7 is shown in Figure 2 (right).

Example 2.9. Consider $w \in W(A_5)$ from Example 2.4. This element has two commutation classes, and hence two heaps as given in Figure 3, where we have color-coded in pink the blocks of each heap that correspond to the braid relation $s_2 s_3 s_2 = s_3 s_2 s_3$. Figure 3 (left) corresponds to the commutation class $\{s_1 s_2 s_3 s_4 s_5 s_2, \ s_1 s_2 s_3 s_4 s_2 s_5, \ s_1 s_2 s_3 s_2 s_4 s_5\}$, while Figure 3 (right) corresponds to the commutation class $\{s_1 s_3 s_2 s_3 s_4 s_5, \ s_3 s_1 s_2 s_3 s_4 s_5\}$.

Given a heap, we can write a reduced expression for the corresponding group element by reading off the generators, starting at the top, moving left to right and then down. The expression we obtain is commutation equivalent to any expression to which the heap corresponds.

The Temperley–Lieb algebra. Given a Coxeter graph Γ, we can form the associative algebra TL(Γ), which we call the Temperley–Lieb algebra of type Γ [Graham
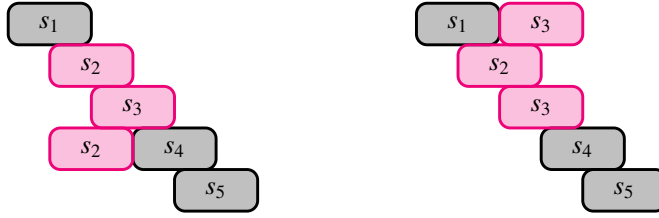
Figure 3. Two different heaps corresponding to the same non-FC element.

1995]. For a complete description of the construction of TL(Γ), see [Ernst 2010; Graham 1995; Green 2006]. For our purposes, it suffices to define TL(A_n) in terms of generators and relations. We are using [Green 2006, Proposition 2.6] (also see [Graham 1995, Proposition 9.5]) as our definition.

The *Temperley–Lieb algebra of type A_n*, denoted by TL(A_n), is the unital $\mathbb{Z}[\delta]$-algebra generated by $\{b_1, b_2, \ldots, b_n\}$ with defining relations

(1) $b_i^2 = \delta b_i$ for all i;

(2) $b_i b_j = b_j b_i$ if $|i - j| > 1$;

(3) $b_i b_j b_i = b_i$ if $|i - j| = 1$.

Suppose w lies in FC(A_n) and has reduced expression $\mathsf{w} = s_{x_1} s_{x_2} \cdots s_{x_k}$. Define the element $b_w \in$ TL(A_n) via

$$b_w = b_{x_1} b_{x_2} \cdots b_{x_k}.$$

Notice that since w is required to be fully commutative, the definition of b_w is independent of choice of reduced expression for w. It is well known (and follows from [Green 2006, Proposition 2.4]) that the set $\{b_w \mid w \in$ FC(A_n)$\}$ forms a $\mathbb{Z}[\delta]$-basis for TL(A_n), called the *monomial basis*.

3. The Temperley–Lieb diagram algebra

Next, we establish our notation and introduce all of the terminology required to define an associative diagram algebra that is a faithful representation of TL(A_n).

Let k be a nonnegative integer. The *standard k-box* is a rectangle with $2k$ points, called *nodes*, labeled as in Figure 4. We will refer to the top of the rectangle as the *north face* and the bottom as the *south face*.

A *concrete pseudo k-diagram* is composed of a finite number of disjoint curves (planar), called *edges*, embedded in the standard k-box with two restrictions:

(1) Every node of the box is the endpoint of exactly one edge, which meets the box transversely.

(2) All other edges must be closed (isotopic to circles) and disjoint from the box.
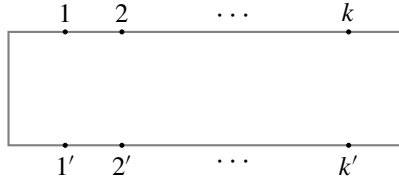
Figure 4. Standard k-box.
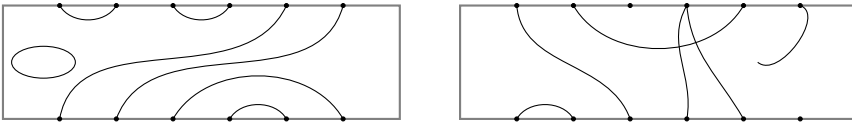
Figure 5. Example of a concrete pseudo 6-diagram together with a nonexample.

Example 3.1. The diagram in Figure 5 (left) is an example of a concrete pseudo 6-diagram, whereas the diagram in Figure 5 (right) does not represent a concrete pseudo 6-diagram since the diagram contains edges that are not disjoint (i.e., they intersect), node 4 is the endpoint for more than one edge, nodes 3 and $6'$ are not endpoints for any edge, and the edge leaving node 6 does not have a node as its second endpoint.

We now define an equivalence relation on the set of concrete pseudo k-diagrams. Two concrete pseudo k-diagrams are (*isotopically*) *equivalent* if one concrete diagram can be obtained from the other by isotopically deforming the edges such that any intermediate diagram is also a concrete pseudo k-diagram. Note that an isotopy of the k-box is a 1-parameter family of homeomorphisms of the k-box to itself that are stationary on the boundary.

A *pseudo k-diagram* is defined to be an equivalence class of equivalent concrete pseudo k-diagrams. We denote the set of pseudo k-diagrams by T_k. Note that we used the word "pseudo" in our definition to emphasize that we allow loops to appear in our diagrams.

Remark 3.2. When representing a pseudo k-diagram with a drawing, we pick an arbitrary concrete representative among a continuum of equivalent choices. When no confusion can arise, we will not make a distinction between a concrete pseudo k-diagram and the equivalence class that it represents. We say that two concrete pseudo k-diagrams are *vertically equivalent* if they are equivalent in the above sense by an isotopy that preserves setwise each vertical cross-section of the k-box.

Example 3.3. The concrete pseudo 5-diagrams in Figure 6 are equivalent since each diagram can be obtained from the other by isotopically deforming the edges.

Figure 6. Isotopically equivalent diagrams.
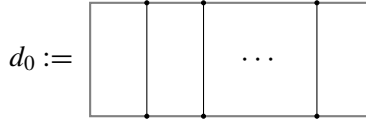
Figure 7. Unique loop-free diagram having only propagating edges.

Let d be a diagram and let e be an edge of d. If e is a closed curve occurring in d, then we call e a *loop*. For example, the diagram in Figure 5 (left) has a single loop. If e joins node i in the north face to node j' in the south face, then e is called a *propagating edge from i to j'*. If e is not propagating, loop or otherwise, it will be called *nonpropagating*. It is clear that there is a unique loop-free diagram consisting only of propagating edges. This diagram, denoted by d_0, is depicted in Figure 7.

We wish to define an associative algebra that has the pseudo k-diagrams as a basis. Let R be a commutative ring with 1. The associative algebra $\mathcal{P}_k$ over R is the free R-module having T_k as a basis. We define multiplication (referred to as diagram concatenation) in $\mathcal{P}_k$ by defining multiplication in the case where d and d' are basis elements, and then extending bilinearly. If $d, d' \in T_k$, the product $d'd$ is the element of T_k obtained by placing d' on top of d, so that node i' of d' coincides with node i of d, and then removing the identified boundary to recover a standard k-box. If desired, one can then vertically rescale the resulting rectangle.

Example 3.4. Figure 8 depicts the product of two pseudo 5-diagrams in $\mathcal{P}_5$.

We now restrict our attention to the base ring $\mathbb{Z}[\delta]$, which is the ring of polynomials in δ with integer coefficients. We define the *Temperley–Lieb diagram algebra* $\mathbb{D}\mathrm{TL}(A_n)$ to be the associative $\mathbb{Z}[\delta]$-algebra equal to the quotient of $\mathcal{P}_{n+1}$ determined by the relation depicted in Figure 9.
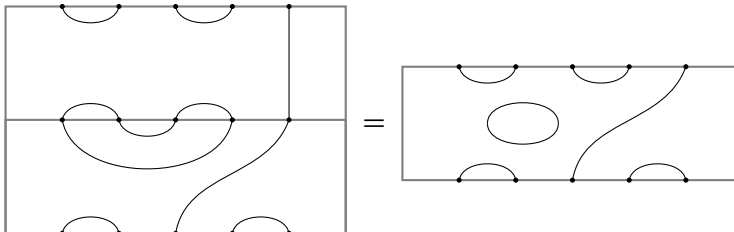
Figure 8. Example of multiplication in $\mathcal{P}_5$.

Figure 9. Defining relation of $\mathbb{D}TL(A_n)$.

It is well known [Kauffman 1987; 1990] that $\mathbb{D}TL(A_n)$ is the free $\mathbb{Z}[\delta]$-module with basis given by the elements of T_{n+1} having no loops. The multiplication is inherited from the multiplication on $\mathcal{P}_{n+1}$ except we multiply by a factor of δ for each resulting loop and then discard the loop. It is easy to see that the identity in $\mathbb{D}TL(A_n)$ is the diagram d_0 given in Figure 7. Technically, the identity diagram is the image of d_0 in the quotient algebra, but there is no danger of identifying the two diagrams.

Example 3.5. Figure 10 depicts the product of three basis diagrams from $\mathbb{D}TL(A_4)$.

Define the *simple diagrams* $d_1, d_2, \ldots, d_n$ as in Figure 11. Note that the simple diagrams are elements of the basis for $\mathbb{D}TL(A_n)$. It turns out [Kauffman 1987; 1990] (and follows from Proposition 3.6 below) that the set of loop-free diagrams of $\mathbb{D}TL(A_n)$ is generated as a unital algebra by the set of simple diagrams $\{d_1, d_2, \ldots, d_n\}$. In fact, $\mathbb{D}TL(A_n)$ is often defined to be the unital $\mathbb{Z}[\delta]$-algebra generated by the simple diagrams subject to the relation given in Figure 9.

It is easy to verify that the simple diagrams of $\mathbb{D}TL(A_n)$ satisfy the defining relations of $TL(A_n)$. That is, we have

(1) $d_i d_i = \delta d_i$ for all i;

(2) $d_i d_j = d_j d_i$ when $|i - j| > 1$;

(3) $d_i d_j d_i = d_i$ when $|i - j| = 1$.

For example, Figure 12 illustrates the third relation above for the case $j = i + 1$. Indeed, $TL(A_n)$ and $\mathbb{D}TL(A_n)$ are isomorphic as $\mathbb{Z}[\delta]$-algebras (for instance, see [Kauffman 1990, §3]).
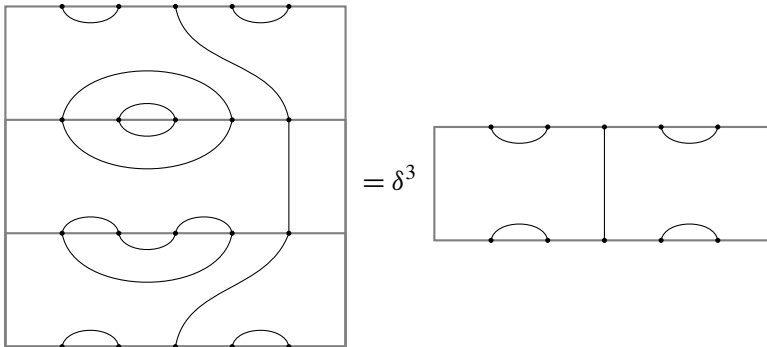
Figure 10. Example of multiplication in $\mathbb{D}TL(A_4)$.
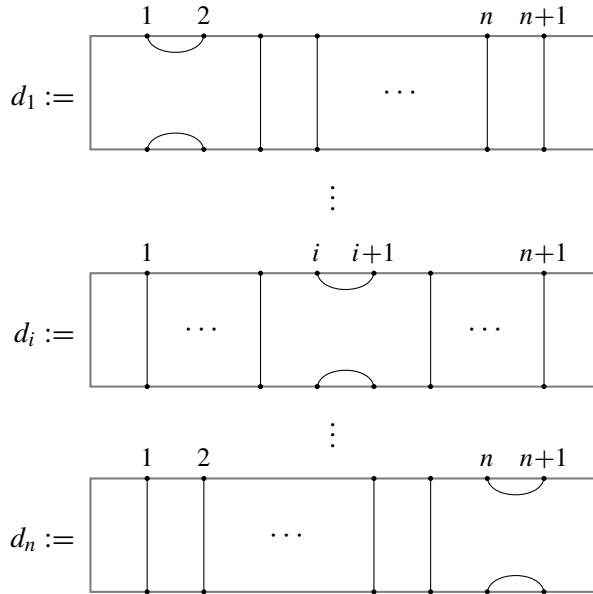
Figure 11. Simple diagrams.
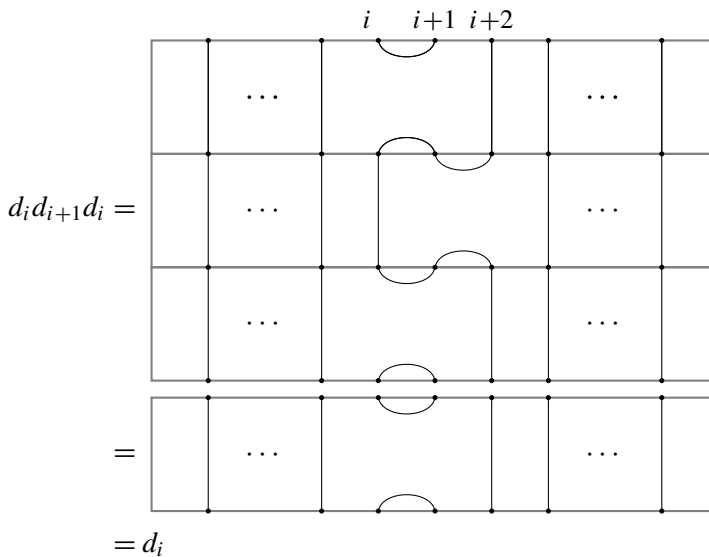
Figure 12. Special case of one of the relations in $\mathbb{D}\mathrm{TL}(A_n)$.

Proposition 3.6. *Let* $\theta : \mathrm{TL}(A_n) \to \mathbb{D}\mathrm{TL}(A_n)$ *be the function determined by* $b_i \mapsto d_i$. *Then* θ *is a well-defined* $\mathbb{Z}[\delta]$-*algebra isomorphism that maps the monomial basis of* $\mathrm{TL}(A_n)$ *to the set of loop-free diagrams in* $\mathbb{D}\mathrm{TL}(A_n)$. $\square$
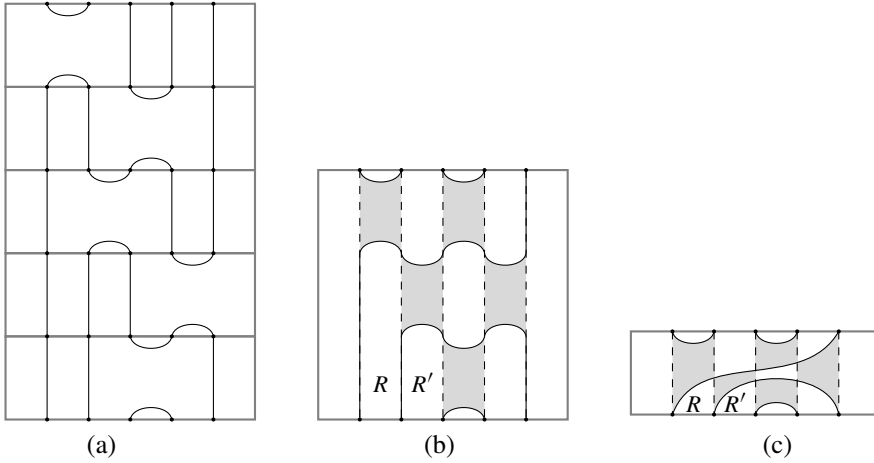
Figure 13. Multiplication of simple diagrams together with the corresponding simple representation and resulting product. The 1-regions of the simple representation and product have been shaded.

Let w be an element of $FC(A_n)$ and define d_w to be the image of the monomial b_w. It follows that if w has $s_{x_1} \cdots s_{x_k}$ as reduced expression, then $d_w = d_{x_1} \cdots d_{x_k}$. That is, given a reduced factorization for $w \in FC(A_n)$, we can easily obtain a reduced factorization of d_w in terms of simple diagrams. However, given a loop-free diagram d, it is more difficult to obtain a factorization. Resolving this difficulty is the content of Section 4.

Let $\mathsf{w} = s_{x_1} \cdots s_{x_k}$ be a reduced expression for $w \in FC(A_n)$. For each simple diagram d_{x_i}, fix a concrete representation such that the propagating edges are straight and the pair of nonpropagating edges never double-back on themselves (i.e., the nonpropagating edges never intersect any vertical line more than once). Now, consider the concrete diagram that results from concatenating the concrete simple diagrams $d_{x_1}, \ldots, d_{x_k}$, rescaling vertically to recover the standard $(n+1)$-box, but not deforming any of the nonpropagating edges. Since w is FC and vertical equivalence respects commutation, given any two reduced expressions for w, the corresponding concrete diagrams constructed as above will be vertically equivalent (see Remark 3.2). We define the corresponding vertical equivalence class to be the *simple representation of d_w*. The simple representation of d_w is designed to replicate the structure of the corresponding heap.

Example 3.7. Let $\mathsf{w} = s_1 s_3 s_2 s_4 s_3$ be a reduced expression for $w \in FC(A_4)$. The factorization for d_w determined by w together with its simple representation is shown in Figure 13(a) and (b), respectively. The resulting product is d_w, which is shown in Figure 13(c). The shaded regions in Figure 13(b) and (c) indicate that the pair of edges bounding the top and bottom of the region arise from the same factor.

In light of Proposition 3.6, it follows that if d is a loop-free diagram from $\mathbb{DTL}(A_n)$, then there exists a unique $w \in \mathrm{FC}(A_n)$ such that $d_w = d$. The upshot is that it makes sense to refer to the simple representation of d.

4. Main results

In this section, we assume that all diagrams are loop-free and that no edge ever double-backs on itself (i.e., other than vertical propagating edges, edges never intersect any vertical line more than once). If d is a k-diagram, we section the corresponding k-box into *columns* by connecting node i in the north face to node i' in the south face. The i-th column C_i lies between nodes i and $i+1$. The connected components of the complement of the edges in each column are called *regions*. For example, the columns and regions for the diagram given in Figure 14(a) are depicted in Figure 14(b).

Lemma 4.1. *The number of edges within a single column of a diagram is even.*

Proof. This is clear for the simple representation of a diagram as each simple diagram d_i contributes precisely two edges to the column C_i. Isotopically deforming

(a) A loop-free diagram d

(b) Regions R, R' are horizontally adjacent

(c) Shaded 1-regions of d

(d) Directed graph G_d

(e) Labeled directed graph G_d^S

(f) Heap corresponding to G_d^S
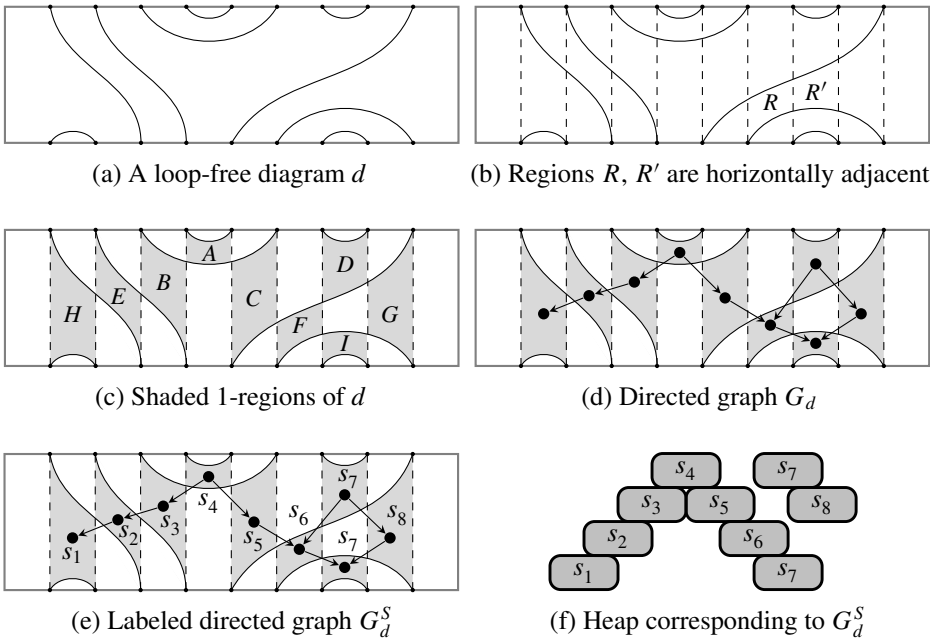
Figure 14. Shaded 1-regions, directed graph G_d, and labeled directed graph G_d^S for a diagram d together with the lattice point representation of the corresponding heap.

edges (while avoiding edges doubling-back on themselves) does not change the parity of the number of edges within the column. □

Lemma 4.2. *The number of regions in each column is odd.*

Proof. Since the number of edges within any column is even, there must be an odd number of regions within each column. □

Note that isotopically deforming the edges of a concrete diagram preserves the relative adjacency of the regions in each column. We say that regions R and R' of column C are *vertically adjacent* if they are adjacent across a common edge. Within a single column, we will label the first region just below the north face with a 0. Moving south, the next region will be labeled with a 1 and we continue this way, alternating labels 0 and 1. We will refer to the labeled regions as 0-regions and 1-regions, respectively. By Lemma 4.2, it is clear that the southernmost region in each column is a 0-region. Figure 14(c) depicts a diagram and its 0-regions and 1-regions, where we have shaded the 1-regions.

Observe that if d is a diagram from $\mathbb{DTL}(A_n)$, then each 1-region in column C_i of the simple representation for d corresponds precisely to the regions bounded above and below by the pair of edges corresponding to a unique factor d_i.

Suppose R and R' are regions of adjacent columns C and C', respectively, of some diagram d. We say that R and R' are *horizontally adjacent* if there exist points p and p' in R and R', respectively, such that the line segment joining p and p' does not cross any edge of d. Loosely speaking, R and R' are horizontally adjacent if they are adjacent across the common vertical boundary of C and C'.

Since we forbid edges from doubling-back on themselves, horizontal adjacency of regions is preserved when isotopically deforming the edges of d. This implies that horizontal adjacency is well-defined.

Figure 14(b) depicts two horizontally adjacent regions, R and R'. However, the regions labeled R and R' in the simple representation depicted in Figure 13(b) may appear at first glance to be horizontally adjacent, but they are not. This is evident by looking at the corresponding regions R and R' in Figure 13(c).

If R is a region of column C in diagram d, then the *depth* of R, depth(R), is defined to be the number of regions in C strictly between the north face of d and R. For example, we have depth$(R) = 1$ and depth$(R') = 2$ for the regions R and R' in Figure 14(b). Note that for any diagram, the northernmost region in each column has depth 0. Moreover, every 1-region has an odd depth while every 0-region has an even depth.

Lemma 4.3. *If R and R' are horizontally adjacent regions of a diagram d, then*

(1) $|\text{depth}(R) - \text{depth}(R')| = 1$, *and*

(2) R *is a* 1-*region if and only if* R' *is a* 0-*region.*

Proof. Induction on depth quickly yields (1), while (2) is an immediate consequence of (1). □

We say that regions R and R' of a diagram d are *diagonally adjacent* if there exists a region S that is vertically adjacent to R and horizontally adjacent to R'. In particular, if S lies below R, then we write $R \to R'$.

Lemma 4.4. *If $R \to R'$ in a diagram d, then R is a 1-region if and only if R' is a 1-region.*

Proof. The result follows immediately from the construction of 0-regions and 1-regions together with Lemma 4.3. □

Example 4.5. In Figure 14(c), we see that

$$A \to B \to E \to H, \quad A \to C \to F \to I, \quad D \to F \to I, \quad \text{and} \quad D \to G \to I.$$

Remark 4.6. If C_i is not the leftmost or rightmost column, edges bounding 1-regions in column C_i must pass into its adjacent columns unless an edge connects directly to a node at the top or bottom of C_i. In the leftmost column, C_1, no edge will pass through to the left. Similarly, in the rightmost column, C_n, no edge will pass through to the right. This implies that if R and R' are both 1-regions in the same column C_i with depth$(R') = $ depth$(R) + 2$ (i.e., R and R' are consecutive 1-regions in C_i with R' below R), then there exist 1-regions T and T' in C_{i-1} and C_{i+1}, respectively, such that $R \to T \to R'$ and $R \to T' \to R'$. Loosely speaking, this determines a local checkerboard pattern of 1-regions, as seen in Figure 14(c).

The checkerboard pattern of 0-regions and 1-regions motivates the following definition. Let d be a diagram having 1-regions $R_1, \ldots, R_n$. Define G_d to be the directed graph having

(1) vertex set $V(G_d) := \{R_1, \ldots, R_n\}$ and

(2) directed edges (R_k, R_l) whenever $R_k \to R_l$.

Since we require the edges of d to not double-back on themselves, it is clear that G_d is independent of choice of concrete representation for d; indeed, isotopically deforming the edges and rescaling the rectangle preserves horizontal and vertical adjacency and so diagonal adjacency is also preserved. In particular, if w indexes d, then we can construct G_d using the simple representation of d_w.

Figure 14(d) shows the directed graph G_d for the diagram d given in Figure 14(a). Observe that directed paths correspond to chains of diagonally adjacent regions.

Next, we will append labels from the generating set of the Coxeter group to the vertices of G_d. Define the vertex labeling function $\nu : V(G_d) \to S$ as follows. If R is a 1-region that lies in column i, then $\nu(R) = s_i$. That is, each region is labeled with the generator of the corresponding column. Now, define G_d^S to be the directed

graph G_d together with the labels on the vertices assigned by v. Figure 14(e) shows the labeled directed graph G_d^S for the diagram d given in Figure 14(a).

Each labeled directed graph G_d^S naturally corresponds to a unique labeled Hasse diagram of a heap for some element in $W(A_n)$. It follows from Remark 4.6 and Proposition 2.6 that this element is FC. Figure 14(f) shows the heap that corresponds to the diagram d given in Figure 14(a). It remains to show that the heap determined by G_d^S corresponds to the group element that indexes the diagram d.

Since diagonal adjacency is preserved when isotopically deforming the edges of d, as in the simple representation, a 1-region R in column C_i is bounded above and below by a pair of edges corresponding to the simple diagram d_i. This region

(a)

(b)

(c)

(d)

(e)
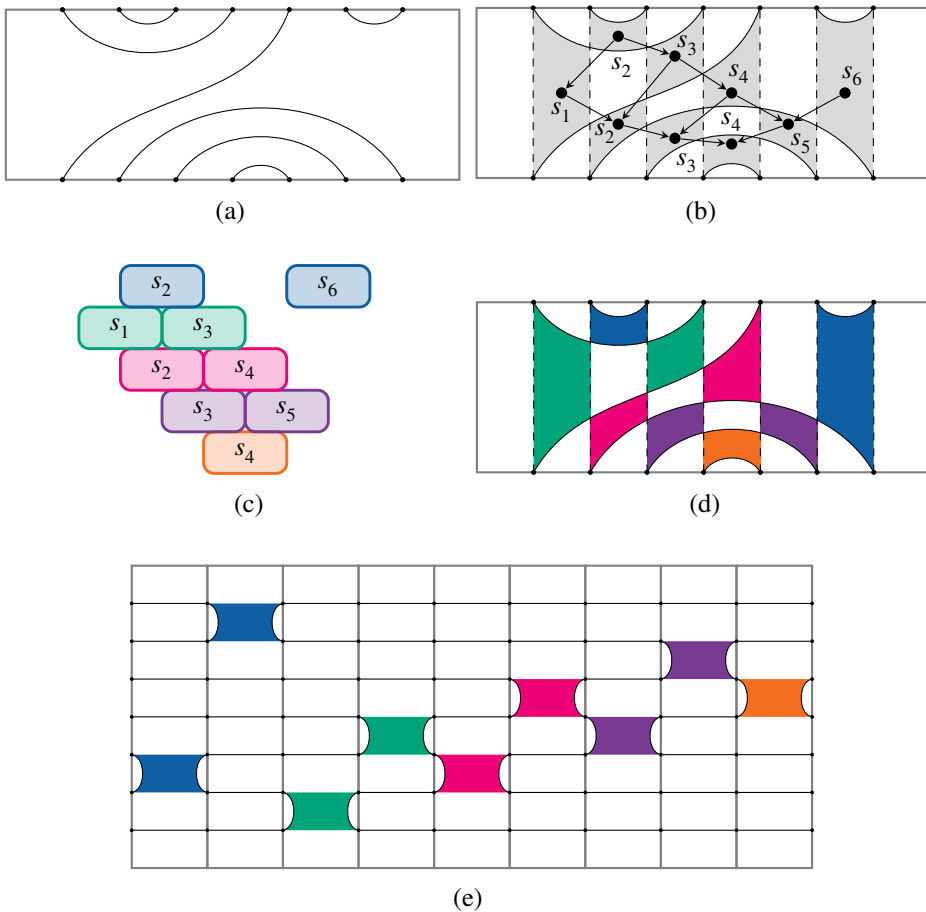
Figure 15. Given a diagram d, we can obtain a reduced factorization by constructing the corresponding labeled directed graph G_d^S, which yields the canonical representation of the heap that indexes d. We have color-coded the corresponding 1-regions and entries of the heap.

is labeled s_i in G_d^S. The structure of G_d^S determines $w \in \text{FC}(A_n)$ satisfying $d = d_w$, and it follows that G_d^S corresponds to $H(w)$. Note that we find w by writing the elements of S corresponding to the labels of each row of the heap left to right starting from the top row and working toward the bottom of the heap.

The above discussion together with the preceding lemmas justifies the following theorem.

Theorem 4.7. *If d is a loop-free diagram in $\text{TL}(A_n)$, then d is indexed by the heap determined by G_d^S.* $\hfill\square$

An immediate consequence of the above theorem is that we nonrecursively obtain a factorization of a diagram d by reading off from top to bottom and left to right the entries of the heap determined by G_d^S. If we choose the canonical representation of the heap, then the factorization of d corresponds to the Cartier–Foata normal form of [Cartier and Foata 1969; Green 2006]. Our construction also yields the following corollary, which appeared independently as Lemma 3.3 in [Green 1998].

Corollary 4.8. *If d is a loop-free diagram in $\text{TL}(A_n)$, then the number of occurrences of the simple diagram d_i in any factorization for d is equal to half the number of edges passing through the column C_i.*

Example 4.9. Consider the diagram d given in Figure 15(a). After forming columns, we obtain a checkerboard of 0-regions and 1-regions, which yields the labeled directed graph G_d^S depicted in Figure 15(b). Then G_d^S determines the canonical representation of the heap given in Figure 15(c), where each row of the heap has a unique color. In Figure 15(d), we have color-coded the 1-regions of d to match the corresponding entries in the heap. By reading off the entries of the heap, we see that $d = d_w$, where

$$w = s_2 s_6 s_1 s_3 s_2 s_4 s_3 s_5 s_4.$$

Equivalently, we obtain the factorization

$$d = d_2 d_6 d_1 d_3 d_2 d_4 d_3 d_5 d_4,$$

which is shown (rotated counterclockwise by a quarter turn in the interest of space) in Figure 15(e).

5. Closing remarks

If (W, S) is a Coxeter system of type Γ, the associated Hecke algebra $\mathcal{H}(\Gamma)$ is an algebra with a basis given by $\{T_w \mid w \in W\}$ and relations that deform the relations of W by a parameter q. As mentioned in Section 1, the ordinary Temperley–Lieb algebra $\text{TL}(A_n)$ is a quotient of the corresponding Hecke algebra $\mathcal{H}(A_n)$. This realization of the Temperley–Lieb algebra as a Hecke algebra quotient was generalized by Graham [1995] to the case of an arbitrary Coxeter system. In

general, the Temperley–Lieb algebra $\mathrm{TL}(\Gamma)$ is a quotient of $\mathcal{H}(\Gamma)$ having several bases indexed by the FC elements of W [Graham 1995, Theorem 6.2].

When a faithful diagrammatic representation of $\mathrm{TL}(\Gamma)$ is known to exist, multiplication in the diagram algebra is given by applying local combinatorial rules to the diagrams. In each case, one can choose a basis for the diagram algebra so that each basis diagram is indexed by an FC element, where the diagrams indexed by the distinguished generators of the Coxeter group form a set of "simple diagrams" that generate the algebra. Every factorization of a basis diagram in terms of simple diagrams corresponds precisely to a factorization of the FC element that indexes the diagram.

Given a reduced expression for an FC element, obtaining the corresponding diagram is straightforward. All one needs to do is concatenate the sequence of simple diagrams determined by the reduced expression and then apply the appropriate local combinatorial rules in the diagram algebra. However, it is another matter to reverse this process. That is, given a basis diagram, can one obtain a factorization in terms of simple diagrams, or equivalently obtain a reduced expression for the FC element that indexes the diagram? Theorem 4.7 answers this question in the affirmative in the case of type A_n. What happens with the other types where faithful diagrammatic representations are known to exist? For example, can we find factorization algorithms for the Temperley–Lieb diagram algebras of types B_n, D_n, E_n, $\widetilde{A}_n$, and $\widetilde{C}_n$?

References

[Billey and Jones 2007] S. C. Billey and B. C. Jones, "Embedded factor patterns for Deodhar elements in Kazhdan–Lusztig theory", *Ann. Comb.* **11**:3–4 (2007), 285–333. MR Zbl

[Cartier and Foata 1969] P. Cartier and D. Foata, *Problèmes combinatoires de commutation et réarrangements*, Lecture Notes in Mathematics **85**, Springer, Berlin, 1969. MR Zbl

[Ernst 2010] D. C. Ernst, "Non-cancellable elements in type affine C Coxeter groups", *Int. Electron. J. Algebra* **8** (2010), 191–218. MR Zbl

[Geck and Pfeiffer 2000] M. Geck and G. Pfeiffer, *Characters of finite Coxeter groups and Iwahori–Hecke algebras*, London Mathematical Society Monographs (N. S.) **21**, Oxford University Press, New York, NY, 2000. MR Zbl

[Graham 1995] J. J. Graham, *Modular representations of Hecke algebras and related algebras*, thesis, University of Sydney, 1995.

[Green 1998] R. M. Green, "Generalized Temperley–Lieb algebras and decorated tangles", *J. Knot Theory Ramifications* **7**:2 (1998), 155–171. MR Zbl

[Green 2006] R. M. Green, "Star reducible Coxeter groups", *Glasg. Math. J.* **48**:3 (2006), 583–609. MR Zbl

[Humphreys 1990] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Studies in Advanced Mathematics **29**, Cambridge University Press, 1990. MR Zbl

[Jones 1999] V. F. R. Jones, "Planar algebras, I", preprint, 1999. arXiv

[Kauffman 1987] L. H. Kauffman, "State models and the Jones polynomial", *Topology* **26**:3 (1987), 395–407. MR Zbl

[Kauffman 1990] L. H. Kauffman, "An invariant of regular isotopy", *Trans. Amer. Math. Soc.* **318**:2 (1990), 417–471. MR Zbl

[Penrose 1971] R. Penrose, "Angular momentum: an approach to combinatorial space-time", pp. 151–180 in *Quantum theory and beyond*, edited by T. Bastin, Cambridge University Press, 1971.

[Stembridge 1996] J. R. Stembridge, "On the fully commutative elements of Coxeter groups", *J. Algebraic Combin.* **5**:4 (1996), 353–385. MR Zbl

[Temperley and Lieb 1971] H. N. V. Temperley and E. H. Lieb, "Relations between the 'percolation' and 'colouring' problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the 'percolation' problem", *Proc. Roy. Soc. London Ser. A* **322**:1549 (1971), 251–280. MR Zbl

[Viennot 1986] G. X. Viennot, "Heaps of pieces, I: Basic definitions and combinatorial lemmas", pp. 321–350 in *Combinatoire énumérative* (Montreal, QC, 1985), edited by G. Labelle and P. Leroux, Lecture Notes in Mathematics **1234**, Springer, Berlin, 1986. MR Zbl

dana.ernst@nau.edu   *Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86011, United States*

mgh64@nau.edu   *Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86011, United States*

sarah.salmon@colorado.edu   *Department of Mathematics, University of Colorado Boulder, Boulder, CO 80309, United States*

Prime labelings of generalized Petersen graphs

Steven A. Schluchter, Justin Z. Schroeder, Kathryn Cokus,
Ryan Ellingson, Hayley Harris, Ethan Rarity and Thomas Wilson

(Communicated by Joseph A. Gallian)

A graph G is called *prime* if the vertices of G can be assigned distinct labels $1, 2, \ldots, |V(G)|$ such that the labels on any two adjacent vertices are relatively prime. By showing that for every even $n \le 2.468 \times 10^9$ there exists $s \in [1, n-1]$ such that both $n + s$ and $2n + s$ are prime, we prove the generalized Peterson graph $P(n, 1)$ is prime for all even $n \in [4, 2.468 \times 10^9]$. Moreover, for a fixed n we describe a method for labeling $P(n, k)$ that is a prime labeling for multiple values of k. Using this method, we prove $P(n, k)$ is prime for all even $n \le 50$ and all odd $k \in [1, n/2)$.

1. Introduction

For a simple graph G with vertex set $V = \{v_1, v_2, \ldots, v_m\}$ and edge set E, a *prime labeling* of G is a bijection $f : V \to \{1, 2, \ldots, m\}$ such that $f(v_i)$ and $f(v_j)$ are relatively prime for all $\{v_i, v_j\} \in E$. A graph is called *prime* if it admits a prime labeling. This concept was proposed by Entringer, who conjectured that all trees are prime, and the first appearance of this problem in print was due to Tout, Dabboucy, and Howalla [Tout et al. 1982]. Since then, many families of graphs have been shown to be prime, including all trees on $m \le 50$ vertices [Pikhurko 2002; 2007] and the grid graph $P_m \times P_n$ when $m \le n$ and n is prime [Sundaram et al. 2006]. More recently, Kh. Md. Mominul Haque, Lin Xiaohui, Yang Yuansheng and Zhao Pingzhong have shown that the generalized Petersen graph $P(n, k)$ is prime for all even $n \le 2500$ when $k = 1$ [Haque et al. 2010] and for all even $n \le 100$ when $k = 3$ [Haque et al. 2011]. Both Diefenderfer et al. [2015] and Prajapati and Gajjar [2014] have shown that $P(n, 1)$ is prime for an infinite family of values of n, the former for $n - 1$ prime and the latter for $n + 1$ prime. Their results are presented as labelings of the prism graph $C_n \times P_2$, which is isomorphic to $P(n, 1)$. For a more thorough treatment of graph labeling, including prime labeling, see [Gallian 2015].
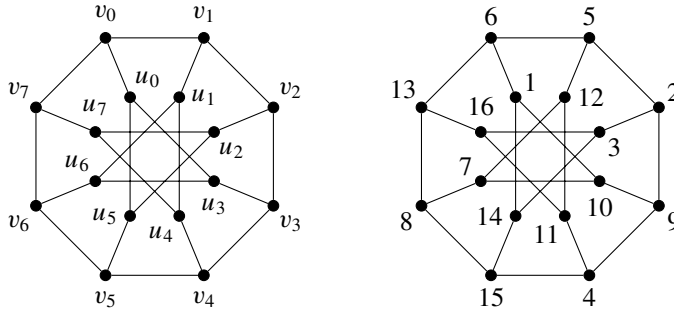
Figure 1. The graph $P(8, 3)$ and a prime labeling of $P(8, 3)$.

We follow the notation of [Haque et al. 2010; 2011]. Since all edges under consideration are undirected, we will write the edge $\{v, w\}$ as vw or wv. For integers $n \geq 3$ and $k \in [1, n/2)$, the *generalized Petersen graph* $P(n, k)$ is defined to be the graph with vertex set $V = \{v_i, u_i : i \in [0, n-1]\}$ and edge set $E = \{v_i v_{i+1}, \ v_i u_i, \ u_i u_{i+k} : i \in [0, n-1]\}$, where all subscripts are reduced modulo n. We will refer to the vertices $v_0, v_1, \ldots, v_{n-1}$ as the *v-vertices* of $P(n, k)$ and the vertices $u_0, u_1, \ldots, u_{n-1}$ as the *u-vertices* of $P(n, k)$. An unlabeled $P(8, 3)$ and a prime labeling of $P(8, 3)$ are shown in Figure 1.

An *independent set* in a graph G is a subset of the vertices of G, no two of which are adjacent. Let $\alpha(G)$ denote the *independence number of G*, i.e., the size of a maximum independent set in G. Given a prime labeling of a graph G with m vertices, the vertices with even labels necessarily form an independent set; thus, $\alpha(G) \geq \lfloor m/2 \rfloor$. It was shown in [Fox et al. 2012] that $\alpha(P(n, k)) < n$ if n is odd or k is even, which leads immediately to the following result (an alternate proof is given in [Prajapati and Gajjar 2015]).

Theorem 1.1. *If n is odd or k is even, then $P(n, k)$ is not prime.*

In this paper, we build on the work appearing in [Diefenderfer et al. 2015; Haque et al. 2010; 2011; Prajapati and Gajjar 2014] by considering prime labelings of $P(n, 1)$ for $n > 2500$ and $P(n, k)$ for small n and $k > 3$. In Section 2, we conjecture that for every even n there exists $s \in [1, n-1]$ such that both $n + s$ and $2n + s$ are prime. In Section 3, we demonstrate a labeling scheme that relies on this conjecture and use computer-generated results to establish that $P(n, 1)$ is prime for all even $n \in [4, 2.468 \times 10^9]$, which improves considerably upon the upper bound given in [Haque et al. 2010]. In Section 4, we fix $n \leq 50$ and describe a method that produces a labeling of $P(n, k)$ that is prime for multiple values of k; with some minor ad hoc switching, this labeling method is used to show $P(n, k)$ is prime for all even $n \in [4, 50]$ and all odd $k \in [1, n/2)$. Together the results of Sections 3 and 4 provide evidence that $P(n, k)$ is prime precisely when n is even and k is odd (as

conjectured in [Prajapati and Gajjar 2015]). Since this is a necessary and sufficient condition for $P(n, k)$ to be bipartite, we reformulate the conjecture as follows.

Conjecture 1.2. $P(n, k)$ *is prime if and only if it is bipartite.*

The property of being bipartite is, in general, neither a necessary nor a sufficient condition for a graph to be prime. The cycle C_n is prime but not bipartite for odd $n \geq 3$, and the complete bipartite graph $K_{n,n}$ is bipartite but not prime for all $n \geq 3$.

For positive integers a and b, we let $\gcd(a, b)$ denote the greatest common divisor of a and b. Then a and b are relatively prime if and only if $\gcd(a, b) = 1$. Note that if $d \mid a$ and $d \mid b$, then $d \mid (a + b)$ and $d \mid (a - b)$. From this we make the following observation.

Lemma 1.3. *Let a and b be positive integers. Then $\gcd(a, b) = 1$ if any of the following hold*:

(1) $a = 1$;

(2) $b = a + 1$;

(3) $a + b$ *is prime*;

(4) $a - b = p$ *is prime and a and b are not multiples of p*;

(5) $a = n + 1$, $b = 2n$ *for some even n.*

2. On the distribution of prime numbers

In a letter to Euler in 1742, Goldbach conjectured that every integer greater than 2 could be written as the sum of three primes (note that he was including 1 as a prime). Euler then reformulated this conjecture in the form in which it is now famous; see [Dickson 2005, pp. 421–424].

Goldbach conjecture. *Every even number greater than 2 can be written as the sum of two primes.*

A closely related conjecture, whose first appearance is due to Maillet [1905], states that every even number can be written as the *difference* of two primes; both this and the Goldbach conjecture remain unsolved. For more information on these conjectures, see [Dickson 2005].

We are going to strengthen Maillet's conjecture by requiring that the two primes be taken from specific intervals. Namely, we conjecture that every even integer n can be written as the difference of primes p_1 and p_2, where $n < p_1 < 2n$ and $2n < p_2 < 3n$. We reformulate this as follows.

Conjecture 2.1. *For every even integer $n \geq 2$, there exists $s \in [1, n - 1]$ such that $n + s$ and $2n + s$ are prime.*

n	s	S_n	n	s	S_n	n	s	S_n	n	s	S_n
2	1	1	52	9	3	102	7	10	152	27	6
4	3	1	54	5	5	104	3	5	154	3	7
6	1	2	56	15	4	106	21	5	156	1	16
8	3	1	58	15	3	108	23	5	158	15	7
10	3	2	60	7	8	110	3	5	160	33	8
12	5	2	62	27	2	112	15	6	162	29	11
14	3	2	64	3	5	114	13	11	164	3	6
16	15	1	66	5	9	116	51	3	166	15	6
18	1	3	68	3	5	118	21	6	168	11	15
20	3	1	70	9	5	120	11	15	170	9	9
22	9	2	72	7	6	122	27	3	172	9	8
24	5	4	74	9	4	124	3	6	174	5	12
26	15	2	76	21	3	126	5	12	176	15	8
28	3	2	78	1	7	128	21	4	178	3	8
30	1	6	80	3	5	130	9	6	180	13	13
32	9	2	82	15	4	132	5	11	182	9	9
34	3	2	84	5	10	134	3	8	184	15	5
36	1	7	86	21	3	136	21	4	186	7	14
38	3	3	88	15	5	138	1	11	188	3	7
40	3	4	90	11	11	140	27	9	190	3	12
42	5	5	92	9	6	142	9	4	192	5	15
44	9	3	94	3	5	144	5	10	194	33	8
46	15	2	96	1	9	146	21	6	196	27	7
48	5	6	98	3	5	148	15	6	198	1	16
50	3	3	100	27	7	150	7	16	200	33	5

Table 1. The minimum value of s such that $n+s$ and $2n+s$ are prime for even $n \in [2, 200]$. Additionally, the number of good s-values S_n is given for each even $n \in [2, 200]$.

For a fixed n, call s *good* if both $n+s$ and $2n+s$ are prime. Table 1 lists good s-values for even $n \in [2, 200]$, and we have obtained good s-values for all even $n \in [2, 2.468 \times 10^9]$ by computer.

Some interesting patterns also occur when we consider S_n, the *number* of good s-values for each n. The first 100 values of S_n are shown in Table 1. Note that $S_n = 1$ for half of the first 10 values of n, but then $S_n > 1$ for all n up to at least 300 million. Moreover, the graph of S_n appears to almost always have a peak at each multiple of 6 and higher peaks at multiples of 30, indicating a strong correlation between S_n and the number of distinct prime factors of n. The portion of this graph from $n = 190,100$ to $n = 190,200$, with straight line segments joining discrete data points, appears in Figure 2 (note that $S_{190,190}$ is a peak that does not occur at a multiple of 6).
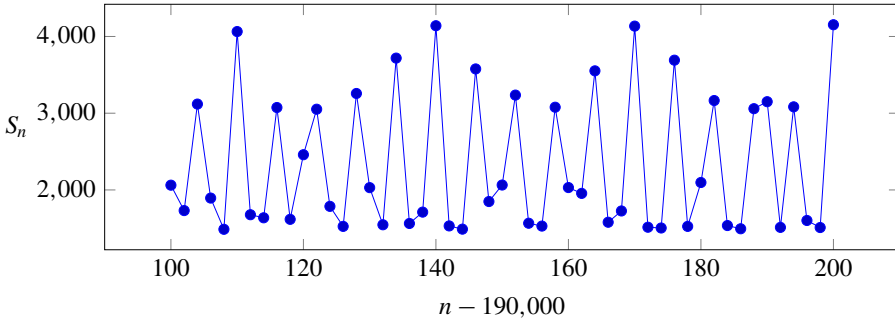
Figure 2. The graph of S_n from $n = 190,100$ to $n = 190,200$.

3. Prime labelings of $P(n, 1)$

The following labeling scheme for $P(n, 1)$ is an expansion and generalization of one employed in [Diefenderfer et al. 2015; Haque et al. 2010; Prajapati and Gajjar 2014].

For any value of k, we know v-vertices with consecutive indices are adjacent. Thus, if we set $f(v_i) = i + 1$ for all $i \in [0, n - 1]$, adjacent v-vertices will either have consecutive integer labels or be labeled with 1 and n; these labels will be relatively prime by Lemma 1.3(2) or (1), respectively. Without loss of generality, we refer to this as a clockwise labeling of the v-vertices (see Figure 3).

If $k = 1$, then u-vertices with consecutive indices are also adjacent, so we label the u-vertices consecutively with $n + 1, n + 2, \ldots, 2n$. Now adjacent u-vertices
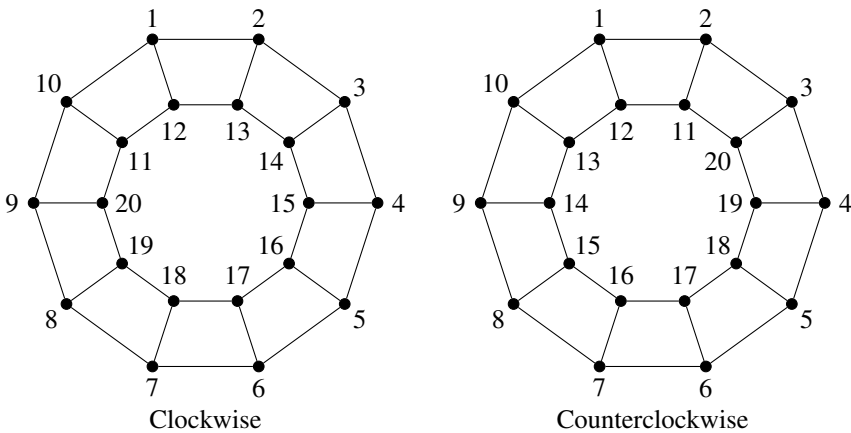
Figure 3. Prime labelings of $P(10, 1)$ using both a clockwise labeling and a counterclockwise labeling for the u-vertices. Note that the *difference* of inner and outer labels in the clockwise labeling is always 1 or 11, and the *sum* of inner and outer labels in the counterclockwise labeling is always 13 or 23.

will either have consecutive integer labels or be labeled with $n+1$ and $2n$; since n is even, these labels will be relatively prime by Lemma 1.3(2) or (5), respectively.

If we apply a clockwise labeling to the u-vertices as well, then labels on each adjacent vu-pair will have a constant difference d (modulo n). By varying our starting point (i.e., where we place the label $n+1$), we can obtain different values of d. The prime labeling of $P(10, 1)$ in Figure 3 employs a clockwise labeling on the u-vertices starting at u_9. For each i the labels on v_i and u_i differ by 1 or 11 ($d \equiv 1 \pmod{10}$), and since $d = 1$ implies consecutive integers and $d = 11$ is prime, these labels are relatively prime by Lemma 1.3(2) or (4), respectively.

If, instead, we apply a counterclockwise labeling to the u-vertices, then labels on each adjacent vu-pair will have a constant sum s (modulo n). Once again, varying the starting point will lead to different values of s. The prime labeling of $P(10, 1)$ in Figure 3 employs a counterclockwise labeling on the u-vertices starting at u_1. For each i the labels on v_i and u_i sum to 13 or 23 ($s \equiv 3 \pmod{10}$), and since both 13 and 23 are prime, these labels are relatively prime by Lemma 1.3(3).

The value of s in the preceding paragraph corresponds to a good s-value from Conjecture 2.1. Let $\mathcal{N}_s = \{n : n + s \text{ is prime}\}$ and $\mathcal{N}_s^* = \{n : 2n + s \text{ is prime}\}$.

Theorem 3.1 [Diefenderfer et al. 2015, Theorem 4.1]. *If $n \in \mathcal{N}_{-1}$, then $P(n, 1)$ is prime.*

Theorem 3.2 [Prajapati and Gajjar 2014, Theorem 2.10]. *If $n \in \mathcal{N}_1$, then $P(n, 1)$ is prime.*

Theorem 3.1 can be obtained by first applying a clockwise labeling to the v-vertices and a clockwise labeling to the u-vertices starting at u_1. Then, switching the labels 1 and $n-1$ on the vertices v_0 and v_{n-2} results in a prime labeling. Note that in [Diefenderfer et al. 2015] the labels n and $2n$ were also switched, but this is not necessary to obtain a prime labeling. Theorem 3.2 was obtained by applying a clockwise labeling to the v-vertices and a clockwise labeling to the u-vertices starting at u_{n-1}.

The following result utilizes a *counterclockwise* labeling on the u-vertices.

Theorem 3.3. *Let $n \geq 4$ be even, and suppose $n \in \mathcal{N}_1^*$ or $n \in \mathcal{N}_s \cap \mathcal{N}_s^*$ for some $s \in [3, n-1]$. Then $P(n, 1)$ is prime.*

Proof. Throughout this proof, we apply a clockwise labeling to the v-vertices starting at v_0 (given by $f(v_i) = i + 1$ for all $i \in [0, n-1]$) and a counterclockwise labeling to the u-vertices starting at some u_j. We have already shown that adjacent v-vertices and adjacent u-vertices will have relatively prime labels, so it remains to consider the labels on v_i and u_i for each i.

If $n \in \mathcal{N}_1^*$, then apply a counterclockwise labeling to the u-vertices starting at u_{n-1}. Formally, let $f(u_i) = 2n - i$ for all $i \in [0, n-1]$. Then the sum of labels

on the edge $v_i u_i$ is $i + 1 + 2n - i = 2n + 1$ for all $i \in [0, n - 1]$. Since $2n + 1$ is prime, these labels are relatively prime by Lemma 1.3(3).

If $n \in \mathcal{N}_s \cap \mathcal{N}_s^*$ for some $s \in [3, n - 1]$, then apply a counterclockwise labeling to the u-vertices starting at u_{s-2}. Formally, let

$$f(u_i) = \begin{cases} n + s - i - 1 & \text{if } i \in [0, s - 2], \\ 2n + s - i - 1 & \text{if } i \in [s - 1, n - 1]. \end{cases}$$

Then the sum of labels on the edge $v_i u_i$ is either $i + 1 + n + s - i - 1 = n + s$ (if $i \in [0, s - 2]$) or $i + 1 + 2n + s - i - 1 = 2n + s$ (if $i \in [s - 1, n - 1]$). Since both $n + s$ and $2n + s$ are prime, these labels are relatively prime by Lemma 1.3(3). $\square$

For an example of the counterclockwise labeling used in Theorem 3.3, see Figure 3. Since $n = 10 \in \mathcal{N}_3 \cap \mathcal{N}_3^*$, we apply a counterclockwise labeling to the u-vertices starting at u_1. The sum of labels on the edge $v_i u_i$ is $n + s = 13$ for $i = 0, 1$ and $2n + s = 23$ for $i \in [2, 9]$.

We have verified Conjecture 2.1 for $n \in [4, 2.468 \times 10^9]$ by computer, and so we have the following result, which further supports the conjecture made in [Haque et al. 2010] that $P(n, 1)$ is prime for all even $n \geq 4$.

Corollary 3.4. $P(n, 1)$ *is prime for all even* $n \in [4, 2.468 \times 10^9]$.

In addition to this bound, note that $\mathcal{N}_{-1} \cup \mathcal{N}_1$ yields an infinite family of generalized Petersen graphs $P(n, 1)$ that are prime.

4. Prime labelings of $P(n, k)$ for even $n \leq 50$ and odd $k > 1$

Here we will give an example of how we produce a labeling of $P(n, k)$ that is a prime labeling for multiple values of $k > 1$. Specifically, we will describe a method of labeling $P(18, k)$ that is prime for $k \in \{3, 5\}$ and show that swapping just two labels also produces a prime labeling of $P(18, 7)$. The tables at the end of this section contain prime labelings of $P(n, k)$ for all even $n \leq 50$ and all odd $k \in [5, n/2)$ obtained using slight variations on the method for $P(18, k)$ (the case $k = 3$ is handled in [Haque et al. 2011]).

If $k > 1$, then u-vertices with consecutive indices are no longer adjacent, so we abandon the clockwise and counterclockwise labeling schemes altogether. Instead, we seek to harness the structure of a bipartition of the vertices of $P(n, k)$. Let $A = \{u_0, v_1, u_2, v_3, \ldots, u_{n-2}, v_{n-1}\}$ and $B = \{v_0, u_1, v_2, u_3, \ldots, v_{n-2}, u_{n-1}\}$ denote the blocks of a bipartition of the vertices of $P(n, k)$; in Figure 4, the top row of vertices forms A while the bottom row forms B. We will place the odd labels on A and the even labels on B in such a way that the resulting labeling is prime for several different values of k.

We begin by placing the even multiples of 3 from left to right in ascending order on the vertices in B with the lowest index (whether v-vertices or u-vertices). We

u_0	v_1	u_2	v_3	u_4	v_5	u_6	v_7	u_8	v_9	u_{10}	v_{11}	u_{12}	v_{13}	u_{14}	v_{15}	u_{16}	v_{17}
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
							9		27			15	3	21		33	

6	12	18	24	30	36												
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
v_0	u_1	v_2	u_3	v_4	u_5	v_6	u_7	v_8	u_9	v_{10}	u_{11}	v_{12}	u_{13}	v_{14}	u_{15}	v_{16}	u_{17}
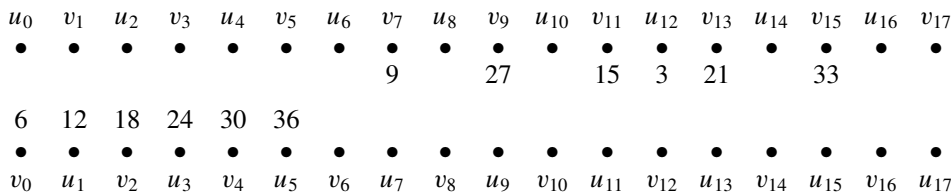
Figure 4. The graph $P(18, k)$ (edges suppressed) labeled with the multiples of 3.

will place the odd multiples of 3 only on v-vertices in A that are not adjacent to any vertex in B labeled with a multiple of 3. However, a simple counting argument shows there will always be one odd multiple of 3 that cannot be placed on a v-vertex; we assume this label is 3 and place the remaining odd multiples of 3 from left to right in ascending order based on the highest prime factor of the label (in the case of a tie, we simply place the smallest integer first—thus 15 is placed before 45, for example). To make this a prime labeling for small values of k, we place 3 on the u-vertex in A whose index is the furthest from the index of any u-vertex in B labeled with a multiple of 3. This first step for $P(18, k)$ is shown in Figure 4.

We continue in a likewise manner for the multiples of 5 and 7, placing the even multiples on the vertices in B with the lowest index and the odd multiples on the available v-vertices in A. If $n \geq 18$, then 35—the only odd multiple of both 5 and 7—is placed on v_{n-1}. Placing these labels may yield adjacent v-vertices that share 5 or 7 as divisors; to correct this, we rearrange the even multiples of 3. In the partial labeling of $P(18, k)$ shown in Figure 5, the labels 25 and 30 on the edge v_3v_4 have a conflict which is fixed in Figure 6 by swapping the labels on v_4 and u_5. It is also possible for larger values of n to have some multiples of 5 or 7 that do not fit on v-vertices; in that case, some additional ad hoc label switching is required.

The remaining even labels with an odd prime factor are placed on the unused vertices in B from left to right in ascending order based on smallest odd prime factor. If there is only one remaining label that is an even multiple of some prime p,

u_0	v_1	u_2	v_3	u_4	v_5	u_6	v_7	u_8	v_9	u_{10}	v_{11}	u_{12}	v_{13}	u_{14}	v_{15}	u_{16}	v_{17}
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	5		25		7		9		27			15	3	21		33		35

6	12	18	24	30	36	10	20	14	28								
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
v_0	u_1	v_2	u_3	v_4	u_5	v_6	u_7	v_8	u_9	v_{10}	u_{11}	v_{12}	u_{13}	v_{14}	u_{15}	v_{16}	u_{17}
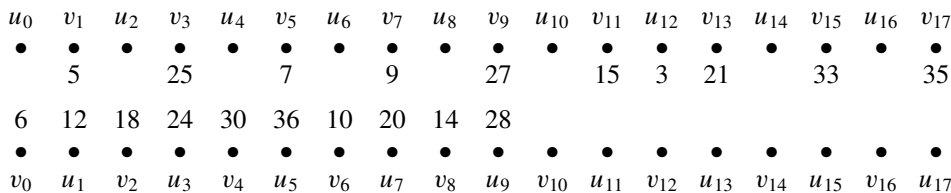
Figure 5. The graph $P(18, k)$ (edges suppressed) labeled with the multiples of 3, 5, and 7. Note that the labels 25 and 30 on the edge v_3v_4 have a conflict, which is addressed in Figure 6.

u_0	v_1	u_2	v_3	u_4	v_5	u_6	v_7	u_8	v_9	u_{10}	v_{11}	u_{12}	v_{13}	u_{14}	v_{15}	u_{16}	v_{17}
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	5		25		7		9		27		15	3	21		33		35
6	12	18	24	**36***	**30***	10	20	14	28	2	22	4	26	8	34	16	32
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
v_0	u_1	v_2	u_3	v_4	u_5	v_6	u_7	v_8	u_9	v_{10}	u_{11}	v_{12}	u_{13}	v_{14}	u_{15}	v_{16}	u_{17}
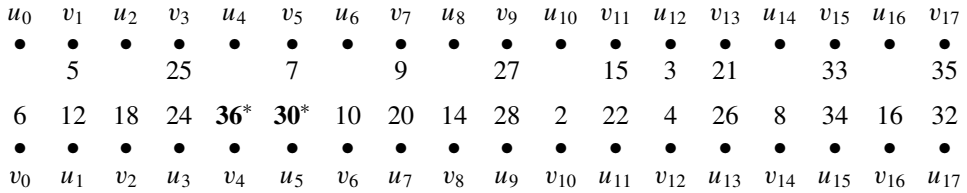
Figure 6. The graph $P(18, k)$ (edges suppressed) labeled with all available even labels and the odd multiples of 3, 5, and 7. Note that the labels on v_4 and u_5 have been switched to fix the conflict on the edge v_3v_4.

u_0	v_1	u_2	v_3	u_4	v_5	u_6	v_7	u_8	v_9	u_{10}	v_{11}	u_{12}	v_{13}	u_{14}	v_{15}	u_{16}	v_{17}
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
1	5	19	25	23	7	**29***	9	31	27	11	15	**3***	21	13	33	17	35
6	12	18	24	36	30	10	20	14	28	2	22	4	26	8	34	16	32
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
v_0	u_1	v_2	u_3	v_4	u_5	v_6	u_7	v_8	u_9	v_{10}	u_{11}	v_{12}	u_{13}	v_{14}	u_{15}	v_{16}	u_{17}
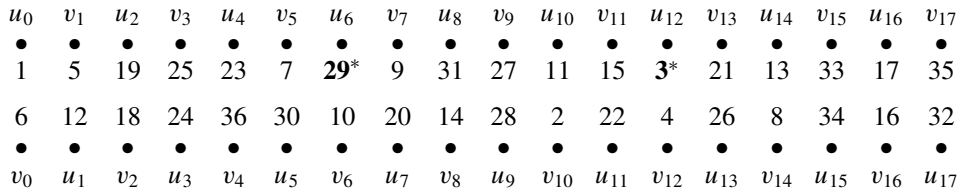
Figure 7. A prime labeling of $P(18, k)$ (edges suppressed) for $k = 3$ or 5. For a prime labeling of $P(18, 7)$, swap the labels on u_6 and u_{12}.

then it is placed on the next available u-vertex. Finally, the multiples of 2 are placed on the unused B vertices, and any necessary label switching from earlier steps is completed. The partial labeling that results for $P(18, k)$ is shown in Figure 6.

Since $2n < 11^2$, the only unused odd multiple of a prime $p > 7$ is p itself; thus, the only odd labels remaining are 1 and the primes greater than 7 and less than $2n$. Each odd prime p less than n is placed on the u-vertex in A whose index is as close as possible to the indices of the u-vertices in B labeled with multiples of p. The remaining labels (1 and the primes greater than n) can be placed arbitrarily on the unlabeled vertices in A. For some n, additional ad hoc label switching may be required. However, for our example $n = 18$, the resulting labeling is prime for $k = 3$ and $k = 5$ (see Figure 7). To obtain a prime labeling for $k = 7$, simply swap the labels on u_6 and u_{12}. For the sake of comparison, the prime labeling of $P(18, 5)$ in Figure 7 is also given in Table 2.

The following result covers the aforementioned ad hoc switching and establishes that, for $n \in [4, 50]$, we have $P(n, k)$ prime precisely when n is even and k is odd (recall from Theorem 1.1 that $P(n, k)$ is *not* prime if n is odd or k is even).

Theorem 4.1. $P(n, k)$ *is prime for all even* $n \in [4, 50]$ *and all odd* $k \in [1, n/2)$.

Proof. The case $k = 1$ was covered in [Haque et al. 2010] and $k = 3$ in [Haque et al. 2011]. Tables 2 and 3 provide a prime labeling of $P(n, k)$ for every even

$P(12,5)$			$P(14,5)$			$P(16,5)$			$P(18,5)$			$P(20,5)$		
i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i
0	6	1	0	6	1	0	6	1	0	6	1	0	6	1
1	5	12	1	5	12	1	5	12	1	5	12	1	5	12
2	18	13	2	18	17	2	18	17	2	18	19	2	18	23
3	7	24	3	7	24	3	19	30	3	25	24	3	25	24
4	10	3	4	10	3	4	24	25	4	36	23	4	36	29
5	9	20	5	9	20	5	7	10	5	7	30	5	7	30
6	14	11	6	14	19	6	20	**23***	6	10	**29***	6	10	**31***
7	15	2	7	27	28	7	9	14	7	9	20	7	9	20
8	22	17	8	22	23	8	28	11	8	14	31	8	40	37
9	21	4	9	15	2	9	27	22	9	27	28	9	27	14
10	8	19	10	26	11	10	2	**3***	10	2	11	10	28	11
11	23	16	11	21	4	11	15	26	11	15	22	11	15	22
			12	8	13	12	4	13	12	4	**3***	12	2	13
			13	25	16	13	21	8	13	21	26	13	21	26
						14	16	29	14	8	13	14	4	**3***
						15	31	32	15	33	34	15	33	34
									16	16	17	16	8	17
									17	35	32	17	39	38
												18	16	19
												19	35	32

Table 2. A prime labeling of $P(n,5)$ for even $n \in [12,22]$. A prime labeling of $P(n,k)$ for even $n \in [16,22]$ and odd $k \in (n/3, n/2)$ can be obtained by swapping the starred labels within a column.

$n \in [12, 42]$ and every odd $k \in [5, n/3)$. If $k > n/3$ is odd (note that n is even, so $n/3$ is also even), then the u-vertex labeled with 3 is adjacent to another u-vertex labeled with a multiple of 3 in the given labeling. Swapping the starred labels in the prime labeling of $P(n,5)$ yields a prime labeling of $P(n,k)$ for even $n \in [16, 42]$ and every odd $k \in (n/3, n/2)$.

Prime labelings of $P(n,k)$ for even $n \in [44, 50]$ and all odd $k \in [5, n/2)$ are given in Tables 4–7. To check these tables, note that for any value of k the labels on all $v_i v_{i+1}$ edges are given by adjacent vertical pairs in the column labeled v_i and the labels on all $v_i u_i$ edges are given by adjacent horizontal pairs. The labels on all $u_i u_{i+k}$ edges are given by vertical pairs at a distance of k in the column labeled u_i. $\qquad\square$

Remark. If n is even one can expand the definition of generalized Petersen graphs to include $P(n, n/2)$. The result is a multigraph with two edges joining u_i and $u_{i+n/2}$ for all $i \in [0, n/2 - 1]$; for prime labeling purposes, these parallel edges can

$P(22,k)$, $k\in[5,7]$			$P(24,k)$, $k\in[5,7]$			$P(26,k)$, $k\in[5,7]$			$P(28,k)$, $k\in[5,9]$			$P(30,k)$, $k\in[5,9]$			$P(32,k)$, $k\in[5,9]$		
i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i
0	6	1	0	42	1	0	6	1	0	6	1	0	6	1	0	6	1
1	5	12	1	5	12	1	5	12	1	5	12	1	5	12	1	5	12
2	18	23	2	18	29	2	18	29	2	18	29	2	18	31	2	18	37
3	25	24	3	25	24	3	25	42	3	25	42	3	25	42	3	25	24
4	30	29	4	48	31	4	48	31	4	54	31	4	54	37	4	42	41
5	7	36	5	35	36	5	7	36	5	55	36	5	55	36	5	55	36
6	42	31	6	6	37	6	24	37	6	24	37	6	24	41	6	48	43
7	37	10	7	7	30	7	49	30	7	7	48	7	7	48	7	49	30
8	20	41*	8	10	41*	8	10	41*	8	30	41	8	30	43	8	54	47
9	9	40	9	9	20	9	9	20	9	49	10	9	49	60	9	7	60
10	14	43	10	40	43	10	40	43	10	20	43*	10	10	47*	10	10	53*
11	27	28	11	27	14	11	27	50	11	9	40	11	9	20	11	9	20
12	26	11	12	28	11	12	14	47	12	50	47	12	40	53	12	40	59
13	15	44	13	15	22	13	15	28	13	27	14	13	27	50	13	27	50
14	22	3*	14	44	13	14	26	11	14	28	53	14	14	59	14	14	61
15	21	2	15	45	26	15	45	44	15	15	56	15	15	28	15	15	28
16	34	13	16	2	3*	16	22	3*	16	26	11	16	56	11	16	56	11
17	33	4	17	21	34	17	21	52	17	45	44	17	45	22	17	45	22
18	38	17	18	4	17	18	2	13	18	22	3*	18	44	13	18	44	13
19	39	8	19	33	38	19	33	34	19	21	52	19	21	26	19	21	26
20	16	19	20	8	19	20	4	17	20	2	13	20	52	3*	20	52	17
21	35	32	21	39	46	21	39	38	21	33	34	21	33	34	21	63	34
			22	16	23	22	8	19	22	4	17	22	2	17	22	2	3*
			23	47	32	23	51	46	23	39	38	23	39	38	23	33	38
						24	16	23	24	8	19	24	4	19	24	4	19
						25	35	32	25	51	46	25	51	46	25	39	46
									26	16	23	26	8	23	26	8	23
									27	35	32	27	57	58	27	51	58
												28	16	29	28	16	29
												29	35	32	29	57	62
															30	32	31
															31	35	64

Table 3. A prime labeling of $P(n,k)$ for even $n\in[24,42]$ and odd $k\in[5,n/3)$. A prime labeling of $P(n,k)$ for even $n\in[24,42]$ and odd $k\in(n/3,n/2)$ can be obtained by swapping the starred labels within a column.

(Continued on next page.)

$P(34,k)$, $k\in[5,11]$			$P(36,k)$, $k\in[5,11]$			$P(38,k)$, $k\in[5,11]$			$P(40,k)$, $k\in[5,13]$			$P(42,k)$, $k\in[5,13]$		
i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i	i	v_i	u_i
0	6	1	0	6	1	0	6	1	0	6	41	0	6	1
1	5	12	1	5	12	1	5	12	1	5	12	1	5	12
2	18	37	2	18	37	2	18	41	2	18	43	2	18	43
3	25	24	3	25	24	3	25	24	3	25	42	3	25	24
4	66	41	4	36	41	4	36	43	4	78	47	4	42	47
5	65	36	5	55	48	5	55	48	5	55	36	5	55	36
6	42	43	6	42	43	6	42	47	6	24	53	6	84	53
7	55	48	7	65	54	7	65	54	7	65	48	7	65	48
8	54	47	8	66	47	8	66	53	8	54	59	8	66	59
9	7	60	9	7	72	9	7	72	9	7	60	9	7	54
10	30	53	10	30	53	10	30	59	10	66	61	10	78	61
11	49	10	11	49	60	11	49	60	11	49	72	11	49	72
12	20	**59***	12	10	**59***	12	10	**61***	12	30	67	12	30	67
13	9	40	13	9	20	13	9	20	13	77	10	13	77	60
14	50	61	14	40	61	14	40	67	14	20	**71***	14	10	**71***
15	27	14	15	27	50	15	27	50	15	9	40	15	9	20
16	28	67	16	70	67	16	70	71	16	50	73	16	40	73
17	15	56	17	39	14	17	39	14	17	27	70	17	27	50
18	26	11	18	28	71	18	28	73	18	2	79	18	70	79
19	45	44	19	45	56	19	45	56	19	15	14	19	81	80
20	22	13	20	26	11	20	26	11	20	28	11	20	14	83
21	21	52	21	21	44	21	75	44	21	45	56	21	15	28
22	34	**3***	22	22	13	22	22	13	22	22	13	22	56	11
23	63	68	23	63	52	23	21	52	23	75	44	23	45	22
24	2	17	24	34	**3***	24	34	**3***	24	26	17	24	44	13
25	33	38	25	33	68	25	63	68	25	21	52	25	75	26
26	4	19	26	38	17	26	38	17	26	34	**3***	26	52	17
27	39	46	27	15	2	27	33	76	27	63	68	27	63	34
28	8	23	28	46	19	28	46	19	28	38	23	28	68	**3***
29	51	58	29	51	4	29	15	2	29	33	76	29	21	38
30	16	29	30	58	23	30	58	23	30	80	19	30	20	19
31	57	62	31	57	8	31	51	4	31	39	46	31	33	46
32	32	31	32	62	29	32	62	29	32	4	29	32	76	23
33	35	64	33	69	16	33	57	8	33	17	58	33	39	58
			34	32	31	34	74	31	34	8	31	34	4	29
			35	35	64	35	69	16	35	57	62	35	51	62
						36	32	37	36	16	37	36	8	31
						37	35	64	37	69	74	37	57	74
									38	32	1	38	16	37
									39	35	64	39	69	82
												40	32	41
												41	35	64

Table 3. (Continued from previous page.)

i	v_i	u_i	departures	i	v_i	u_i	departures
0	6	1		22	56	11	
1	5	12		23	45	22	
2	18	47		24	44	13	
3	55	84		25	75	88	
4	78	53		26	26	17	
5	85	36		27	21	52	
6	42	59		28	34	3	$u_{28}=73$ for $k \in [15, 21]$
7	65	48		29	63	68	
8	54	61		30	38	9	$u_{30}=77$ for $k=15$,
9	25	72					$u_{30}=43$ for $k \in [17, 21]$
10	66	67		31	33	76	
11	7	60		32	46	19	
12	30	71		33	39	2	
13	49	24		34	4	29	
14	10	73	$u_{14}=3$ for $k \in [15, 21]$	35	51	58	
15	77	20	$v_{15}=9$ for $k=15$	36	8	31	
16	40	79		37	57	62	
17	27	50		38	16	37	
18	70	83		39	69	74	
19	81	80		40	32	41	
20	14	23		41	87	82	
21	15	28		42	64	43	$u_{42}=9$ for $k \in [17, 21]$
				43	35	86	

Table 4. A prime labeling of $P(44, k)$ for all odd $k \in [15, 21]$.

be suppressed to a single edge. When $n \in [14, 50]$ and $n/2$ is odd, the prime labeling of $P(n, n/2 - 2)$ given in Theorem 4.1 is also a prime labeling of $P(n, n/2)$. It is a simple exercise to show that $P(6, 3)$ and $P(10, 5)$ are prime, and additional results concerning prime labelings of $P(n, n/2)$ can be found in [Prajapati and Gajjar 2015].

Acknowledgments

References

[Dickson 2005] L. E. Dickson, *History of the theory of numbers, Vol. I: Divisibility and primality*, Dover, New York, 2005. Zbl

i	v_i	u_i	departures	i	v_i	u_i	departures
0	6	1		22	28	11	
1	65	12		23	15	56	
2	18	47		24	22	13	
3	55	24		25	45	44	
4	84	53		26	88	17	
5	85	36		27	75	26	
6	42	59		28	52	3	$v_{28}=4, u_{28}=91$ for $k \in [15, 21]$
7	5	48		29	21	34	
8	78	61		30	68	9	$u_{30}=43$ for $k \in [17, 21]$
9	25	54		31	63	38	
10	66	67		32	76	23	
11	7	72		33	33	46	
12	30	71		34	92	19	
13	49	60		35	39	58	
14	90	73		36	2	29	
15	77	10		37	51	62	
16	20	79		38	4	31	$v_{38}=52$ for $k \in [15, 21]$
17	91	40	$v_{17}=3$ for $k \in [15, 21]$	39	57	74	
18	50	83		40	8	37	
19	27	70		41	69	82	
20	80	89		42	16	41	
21	81	14		43	87	86	
				44	32	43	$u_{44}=9$ for $k \in [17, 21]$

Table 5. A prime labeling of $P(46, k)$ for all odd $k \in [15, 21]$.

[Diefenderfer et al. 2015] N. Diefenderfer, D. C. Ernst, M. Hastings, L. N. Heath, H. Prawzinsky, B. Preston, J. Rushall, E. White, and A. Whittemore, "Prime Vertex Labelings of Several Families of Graphs", preprint, 2015. To appear in *Involve*. arXiv

[Fox et al. 2012] J. Fox, R. Gera, and P. Stănică, "The independence number for the generalized Petersen graphs", *Ars Combin.* **103** (2012), 439–451. MR Zbl

[Gallian 2015] J. A. Gallian, "A dynamic survey of graph labeling", *Electron. J. Combin.* **5** (2015), Dynamic Survey 6, 389 pp.

[Haque et al. 2010] K. M. M. Haque, X. Lin, Y. Yang, and P. Zhao, "On the prime labeling of generalized Petersen graph $P(n, 1)$", *Util. Math.* **83** (2010), 95–106. MR Zbl

[Haque et al. 2011] K. M. M. Haque, X. Lin, Y. Yang, and P. Zhao, "On the prime labeling of generalized Petersen graphs $P(n, 3)$", *Int. J. Contemp. Math. Sci.* **6**:33–36 (2011), 1783–1800. MR Zbl

[Maillet 1905] E. Maillet, "Réponse à la question 574", *L'Intermédiaire des mathématiciens* **12** (1905), 107–109.

[Pikhurko 2002] O. Pikhurko, "Every tree with at most 34 vertices is prime", *Util. Math.* **62** (2002), 185–190. MR Zbl

i	v_i	u_i	departures	i	v_i	u_i	departures
0	6	43		24	56	11	
1	55	12		25	45	22	
2	18	47		26	44	13	
3	65	24		27	75	88	
4	84	1		28	26	17	
5	85	36		29	21	52	
6	42	53		30	34	27	$u_{30}=77$ for $k \in [15, 23]$
7	95	48		31	63	68	
8	54	59		32	38	3	$u_{32}=73$ for $k \in [17, 23]$
9	5	96		33	33	76	
10	66	61		34	46	9	$u_{34}=91$ for $k \in [15, 23]$
11	25	72		35	39	92	
12	78	67		36	2	19	
13	7	30		37	51	58	
14	90	71		38	4	23	
15	49	60		39	57	62	
16	10	73	$u_{16}=3$ for $k \in [17, 23]$	40	8	29	
17	77	20	$v_{17}=27$ for $k \in [15, 23]$	41	69	74	
18	40	79		42	16	31	
19	91	50	$v_{19}=9$ for $k \in [15, 23]$	43	87	82	
20	80	83		44	32	37	
21	81	70		45	93	86	
22	14	89		46	64	41	
23	15	28		47	35	94	

Table 6. A prime labeling of $P(48, k)$ for all odd $k \in [15, 23]$.

[Pikhurko 2007] O. Pikhurko, "Trees are almost prime", *Discrete Math.* **307**:11–12 (2007), 1455–1462. MR Zbl

[Prajapati and Gajjar 2014] U. M. Prajapati and S. J. Gajjar, "Some results on prime labeling", *Open J. Discrete Math.* **4**:3 (2014), 60–66.

[Prajapati and Gajjar 2015] U. M. Prajapati and S. J. Gajjar, "Prime labeling of generalized Petersen graph", *Int. J. Math. and Soft Comp.* **5**:1 (2015), 65–71.

[Sundaram et al. 2006] M. Sundaram, R. Ponraj, and S. Somasundaram, "On a prime labeling conjecture", *Ars Combin.* **79** (2006), 205–209. MR Zbl

[Tout et al. 1982] A. Tout, A. N. Dabboucy, and K. Howalla, "Prime labeling of graphs", *Nat. Acad. Sci. Letters* **11** (1982), 365–368.

sschluch@gmu.edu   *Department of Mathematical Sciences,*
                                    George Mason University, 4400 University Drive, MS: 3F2,
                                    Fairfax, VA 22030, United States

i	v_i	u_i	departures	i	v_i	u_i	departures
0	6	43		25	45	56	
1	5	12		26	98	11	
2	18	47		27	75	22	
3	25	24		28	44	13	
4	84	1		29	21	88	
5	55	36		30	26	9	$u_{30} = 77$ for $k \in [15, 23]$
6	42	53		31	63	52	
7	65	48		32	34	3	$u_{32} = 73$ for $k \in [17, 23]$
8	54	59		33	15	68	
9	85	96		34	38	27	$u_{34} = 91$ for $k \in [17, 23]$
10	66	61		35	99	76	
11	95	72		36	46	17	
12	78	67		37	39	92	
13	7	30		38	2	19	
14	90	71		39	51	58	
15	49	60		40	4	23	
16	10	73	$u_{16} = 3$ for $k \in [17, 23]$	41	57	62	
17	77	20	$v_{17} = 27$ for $k \in [15, 23]$	42	8	29	
18	40	79		43	69	74	
19	91	50	$v_{19} = 9$ for $k \in [15, 23]$	44	16	31	
20	80	83		45	87	82	
21	81	70		46	32	37	
22	100	89		47	93	86	
23	33	14		48	64	41	
24	28	97		49	35	94	

Table 7. A prime labeling of $P(50, k)$ for all odd $k \in [15, 23]$.

justin@radstock.org   *Mosaic Centre Radstock, Kej Bratstvo Edinstvo 45,*
                             1230 Gostivar, Macedonia

kcokus@gmu.edu   *George Mason University, 4400 University Drive,*
                             Fairfax, VA 22030, United States

rellings@gmu.edu   *George Mason University, 4400 University Drive,*
                             Fairfax, VA 22030, United States

hharri15@gmu.edu   *George Mason University, 4400 University Drive,*
                             Fairfax, VA 22030, United States

erarity@gmu.edu   *George Mason University, 4400 University Drive,*
                             Fairfax, VA 22030, United States

twilso19@gmu.edu   *George Mason University, 4400 University Drive,*
                             Fairfax, VA 22030, United States

A generalization of Zeckendorf's theorem
via circumscribed m-gons

Robert Dorward, Pari L. Ford, Eva Fourakis, Pamela E. Harris,
Steven J. Miller, Eyvindur Palsson and Hannah Paugh

(Communicated by Stephan Garcia)

Zeckendorf's theorem states that every positive integer can be uniquely decomposed as a sum of nonconsecutive Fibonacci numbers, where the Fibonacci numbers satisfy $F_1 = 1$, $F_2 = 2$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 3$. The distribution of the number of summands in such a decomposition converges to a Gaussian, the gaps between summands converge to geometric decay, and the distribution of the longest gap is similar to that of the longest run of heads in a biased coin; these results also hold more generally, though for technical reasons previous work is needed to assume the coefficients in the recurrence relation are nonnegative and the first term is positive.

We extend these results by creating an infinite family of integer sequences called the m-gonal sequences arising from a geometric construction using circumscribed m-gons. They satisfy a recurrence where the first $m+1$ leading terms vanish, and thus cannot be handled by existing techniques. We provide a notion of a legal decomposition, and prove that the decompositions exist and are unique. We then examine the distribution of the number of summands used in the decompositions and prove that it displays Gaussian behavior. There is geometric decay in the distribution of gaps, both for gaps taken from all integers in an interval and almost surely in distribution for the individual gap measures associated to each integer in the interval. We end by proving that the distribution of the longest gap between summands is strongly concentrated about its mean, behaving similarly as in the longest run of heads in tosses of a coin.

1. Introduction

The Fibonacci numbers are a heavily studied sequence which arises in many different ways and places. By defining them as $F_1 = 1$, $F_2 = 2$ and $F_{n+1} = F_n + F_{n-1}$, we have the remarkable property that every positive integer can be uniquely written as a sum of nonconsecutive Fibonacci numbers; further, this property is equivalent

---

to the definition of the Fibonacci numbers (i.e., if $\{a_n\}$ is a sequence of numbers such that every integer can be written uniquely as a sum of nonadjacent terms in the sequence, then $\{a_n\} = \{F_n\}$). Zeckendorf [1972] proved this in 1939, though he did not publish this result until much later.

In recent years many have studied generalizations to Zeckendorf's theorem by exploring different notions of decompositions and the properties of the associated sequences; see among others [Alpert 2009; Daykin 1960; Demontigny et al. 2014a; 2014b; Drmota and Gajdosik 1998; Filipponi et al. 1994; Grabner and Tichy 1990; Grabner et al. 1994; Keller 1972; Lengyel 2006; Miller and Wang 2012; 2014; Steiner 2002; 2005]. Despite the vast literature in this area, the majority of the research on generalized Zeckendorf decompositions has involved sequences with *positive linear recurrences*. Positive linear recurrence sequences $\{G_n\}$ satisfy a linear recurrence relation where the coefficients are nonnegative with the first and last term coefficients being positive.[1]

There has been little research which considers cases where the leading coefficient in the recurrence is zero; one such case is found in [Catral et al. 2014]. They studied what they call the Kentucky sequence, which is defined by the recurrence relation $H_{n+1} = H_{n-1} + 2H_{n-3}$ and $H_i = i$ for $i \leq 4$. While the behavior there is similar to the positive linear recurrences, there are sequences with very different behavior. One such is the Fibonacci quilt, which arises from creating a decomposition rule from the Fibonacci spiral[2] (see [Catral et al. 2016a; 2016b]), where the number of decompositions is not unique but in fact grows exponentially. This leads to the major motivation of this paper (as well as the motivation for the three papers just mentioned): how important is the assumption that the leading term is positive? The work mentioned above shows that it is not just a technically convenient assumption; markedly different behavior can emerge. Our goal is to try and determine when we have each type of behavior, and thus the purpose of this paper is to explore infinitely many recurrences with absent leading term and see the effect that has on the properties of the decompositions.

Specifically, we consider an infinite family of integer sequences called the *m-gonal sequences*, where $m \geq 3$. These sequences arise from a geometric construction using circumscribed m-gons, and after defining them below we state our results.

Definition of the m-gonal sequence.

One interpretation of Zeckendorf's theorem, which states that every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers, is that we have infinitely many bins with just one

---

[1] Thus $G_{n+1} = c_1 G_n + \cdots + c_L G_{n-(L-1)}$ with $c_1 c_L > 0$ and $c_i \geq 0$.

[2] Let $f_n = F_{n-1}$, the standard definition of the Fibonacci numbers. Then the plane can be tiled in a spiral where the dimensions of the n-th square are $f_n \times f_n$; we declare a decomposition legal if no two summands used share an edge.
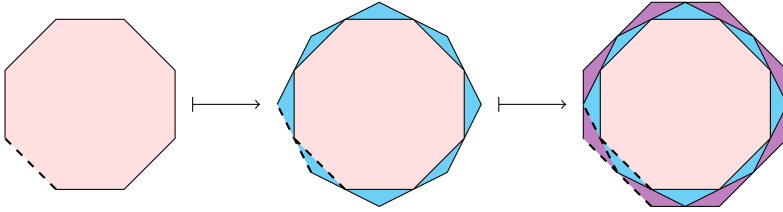
Figure 1. Circumscribed m-gons.

number per bin, and if we choose a bin to contribute a summand to a number's decomposition then we cannot choose a summand from an adjacent bin. We can generalize to bins with more elements, as well as disallow two bins to be used if they are within a given distance (see [Catral et al. 2014; 2016a; 2016b]). The m-gonal sequences are similar to these constructions, but have a two-dimensional structure arising from circumscribing m-gons about one central m-gon.

Briefly we view the decomposition rule corresponding to the m-gonal sequence for $m \geq 1$ by saying the sequence is partitioned into bins b_i of length $|b_i|$, where $|b_0| = 1$ and $|b_i| = m$ for all $i \geq 1$. A valid decomposition has no two summands being elements from the same bin. We refer to this decomposition as a *legal m-gonal decomposition of a positive integer z*. We now give details and examples of this construction.

The m-gonal sequence was initially constructed by circumscribing m-gons. For $m \geq 3$ we let M_0 denote a regular m-gon. Circumscribe the m-gon M_1 onto M_0 such that the vertices of M_0 bisect the edges of M_1. Note that this adds m faces to the resulting figure. We continue this process indefinitely, where we circumscribe the m-gon M_i onto M_{i-1} such that the vertices of M_{i-1} bisect the edges of M_i. At each step we have added an additional m faces to the resulting figure. We depict these initial iterations in Figure 1. Let $\mathcal{M}$ denote all of the faces created through the process of circumscribing m-gons. Then

$$\mathcal{M} = \{f_0\} \cup \left(\bigcup_{i=1}^{\infty} \{f(i, 1), f(i, 2), \ldots, f(i, m)\} \right), \tag{1-1}$$

where f_0 is the face of the m-gon M_0 and $f(i, 1), f(i, 2), \ldots, f(i, m)$ are the faces added to $\mathcal{M}$ when M_i is circumscribed onto M_{i-1} for $i \geq 1$.

Fix an integer $m \geq 3$. Suppose $\{a_n\}_{n=0}^{\infty}$ is an increasing sequence of positive integers. We define the ordered lists $b_0 = [a_0]$ and $b_i = [a_{m(i-1)+1}, a_{m(i-1)+2}, \ldots, a_{mi}]$ for $i \leq 1$, which we refer to as bins. For all $i \geq 1$, we see that b_i has size m and b_0 has size 1. The integers in bin b_i correspond directly with the integers which we place on the faces added to $\mathcal{M}$ when M_i is circumscribed onto M_{i-1}. With the elements of our sequence partitioned into bins, we define a *legal m-gonal decomposition* of

any positive integer z. If we have

$$z = a_{\ell_t} + a_{\ell_{t-1}} + \cdots + a_{\ell_2} + a_{\ell_1}, \tag{1-2}$$

where $\ell_1 < \ell_2 < \cdots < \ell_t$ and $\{a_{\ell_j}, a_{\ell_{j+1}}\} \not\subset b_i$ for any $i \geq 0$ and $1 \leq j \leq t - 1$, then we call this a legal m-gonal decomposition of z. Namely, a legal m-gonal decomposition cannot use more than one summand from the same bin. With the generalized construction of the sequence by partitioning the members into bins rather than relying solely on the 2-dimensional circumscribed polygons, we make it a formal definition for $m \geq 1$.

Definition 1.1. Let an increasing sequence of positive integers $\{a_n\}_{n=0}^{\infty}$ be given and partition the elements into ordered lists

$$b_0 = [a_0], \quad b_k := [a_{m(k-1)+1}, a_{m(k-1)+2}, \ldots, a_{mk}] \tag{1-3}$$

for $m \geq 1$, $k \geq 1$, which we call bins. We declare a decomposition of an integer

$$z = a_{\ell_t} + a_{\ell_{t-1}} + \cdots + a_{\ell_1}, \tag{1-4}$$

where $\ell_1 < \ell_2 < \cdots < \ell_t$ and $\{a_{\ell_j}, a_{\ell_{j+1}}\} \not\subset b_i$ for any i, j, to be a *legal m-gonal decomposition*.

The following definition details the construction of the m-gonal sequence, which is the focus of this paper.

Definition 1.2. For $m \geq 1$, an increasing sequence of positive integers $\{a_n\}_{n=0}^{\infty}$ is called an *m-gonal sequence* if every a_i ($i \geq 0$) is the smallest positive integer that does not have a legal m-gonal decomposition using the elements $\{a_0, a_1, \ldots, a_{i-1}\}$.

Example 1.3. For $m = 1$, all the bins have size 1 and the 1-gonal sequence $\{a_i\}_{i=0}^{\infty}$ is defined by $a_i = 2^i$. This is equivalent to writing an integer in binary. When $m = 2$ we have bins $b_i = [a_{2i-1}, a_{2i}]$ for $i \geq 1$ and $b_0 = [a_0]$. The first few terms of the sequence are

$$\underbrace{1}_{b_0}, \underbrace{2, 4}_{b_1}, \underbrace{6, 12}_{b_2}, \underbrace{18, 36}_{b_3}, \underbrace{54, 108}_{b_4}, \underbrace{162, 324}_{b_5}, \ldots.$$

In the case where $m = 3$ the triangle (3-gonal) sequence begins with the terms

$$\underbrace{1}_{b_0}, \underbrace{2, 4, 6}_{b_1}, \underbrace{8, 16, 24}_{b_2}, \underbrace{32, 64, 96}_{b_3}, \underbrace{128, 256, 384}_{b_4}, \underbrace{512, 1024, 1536}_{b_5}, \ldots.$$

Figure 2 gives a visualization of the beginning of the triangle sequence when the integers are placed in the faces of the circumscribed triangles. Moreover, we note that the triangles used need not be equilateral.

Also one can observe that the triangle decomposition of 2015 is given by

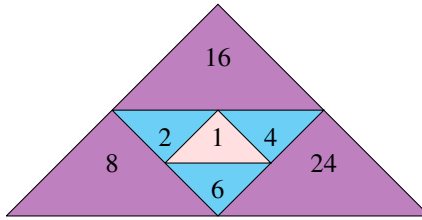$$2015 = a_{15} + a_{12} + a_6 + a_3 + a_0 = 1536 + 384 + 64 + 24 + 6 + 1. \tag{1-5}$$

Figure 2. Beginning of triangle sequence.

In Section 2 we derive the recurrence relation and explicit closed form expressions for the terms of the m-gonal sequence, which we state below.

Theorem 1.4. *Let $m \geq 1$. If $\{a_n\}_{n=0}^{\infty}$ is the m-gonal sequence, then*

$$a_n = \begin{cases} 1 & \text{if } n = 0, \\ 2n & \text{if } 1 \leq n \leq m, \\ (m+1)a_{n-m} & \text{if } n > m. \end{cases} \tag{1-6}$$

Then for $n \geq 1$, with $n = km + r$, $k \geq 0$ and $1 \leq r \leq m$,

$$a_n = 2r(m+1)^k. \tag{1-7}$$

Uniqueness of decomposition. Notice that for $m \geq 2$, the recurrence given in Theorem 1.4 is not a positive linear recurrence, as the leading coefficients of the first m terms are zero. Therefore past results on positive linear recurrences do not apply to the m-gonal sequence; however, we do still obtain unique decomposition.

Theorem 1.5 (uniqueness of decompositions). *Fix $m \geq 1$. Every positive integer can be written uniquely as a sum of distinct terms from the m-gonal sequence, where no two summands are in the same bin.*

A proof of Theorem 1.5 is given in the Appendix.

Gaussianity. Previous work with positive linear recurrence sequences proved the number of summands in the decomposition of positive integers converges to a Gaussian (see among others [Demontigny et al. 2014a; Miller and Wang 2014]). The same holds for Kentucky decompositions, despite the fact that the Kentucky sequence is not a positive linear recurrence [Catral et al. 2014], and also for the m-gonal sequences.

Theorem 1.6 (Gaussian behavior of summands). *Let the random variable Y_n denote the number of summands in the (unique) m-gonal decomposition of an integer picked at random from $[0, a_{mn+1})$ with uniform probability.[3] Normalize Y_n to*

---

[3]Using the methods of [Best et al. 2016], these results can be extended to hold almost surely for a sufficiently large subinterval of $[0, a_{mn+1})$.

$Y'_n = (Y_n - \mu_n)/\sigma_n$, where μ_n and σ_n are the mean and variance of Y_n respectively. Then

$$\mu_n = \frac{mn}{m+1} + \frac{1}{2}, \quad \sigma_n^2 = \frac{mn}{(m+1)^2} + \frac{1}{4}, \tag{1-8}$$

and Y'_n converges in distribution to the standard normal distribution as $n \to \infty$.

The proof of Theorem 1.6 is given in Section 3.

Gaps between summands. Another property of positive linear recurrence sequences that is studied is the behavior of the gaps between adjacent summands in decompositions. In many instances, it has been shown that there is exponential decay in the distribution of gaps; see [Beckwith et al. 2013; Ben-Ari and Miller 2014; Bower et al. 2015].[4] Similarly, the Kentucky sequence displays exponential decay in the distribution of gaps [Catral et al. 2014]. We obtain similar behavior again, though now there is a slight dependence on the residue of gap modulo m (if we split by residue we obtain geometric decay).

Before stating our result we first fix some notation. For the legal m-gonal decomposition

$$z = a_{\ell_k} + a_{\ell_{k-1}} + \cdots + a_{\ell_1} \quad \text{with } \ell_1 < \ell_2 < \cdots < \ell_k \tag{1-9}$$

and $z \in [0, a_{mn+1})$, we define the multiset of gaps as

$$\text{Gaps}_n(z) := \{\ell_2 - \ell_1, \ \ell_3 - \ell_2, \ \ldots, \ \ell_k - \ell_{k-1}\}. \tag{1-10}$$

Observe that we do not consider $\ell_1 - 0$ as a gap. However, doing so would not affect the limiting behavior. For example, notice $z = a_{15} + a_{12} + a_6 + a_3 + a_0$ contributes three gaps of length 3, and one gap of length 6.

Considering all the gaps between summands in legal m-gonal decompositions of all $z \in [0, a_{mn+1})$, we let $P_n(g)$ be the fraction of all these gaps that are of length g. That is, $P_n(g)$ is the probability of a gap of length g among legal m-gonal decompositions of $z \in [0, a_{mn+1})$.

Theorem 1.7 (average gap measure). *Let $g = m\alpha + \beta$, where $\alpha \geq 0$ and $0 \leq \beta < m$. For $P_n(g)$ as defined above, the limit $P(g) := \lim_{n\to\infty} P_n(g)$ exists, and*

$$P(g) = \begin{cases} \frac{\beta}{m(m+1)} & \text{if } \alpha = 0, \\ \frac{m+1-\beta}{(m+1)^{\alpha+1}} & \text{if } \alpha > 0. \end{cases} \tag{1-11}$$

The proof for Theorem 1.7 is given in Section 4.

Via an application of [Dorward et al. 2015, Theorem 1.1] we extract a result on individual gaps for the m-gonal case. In order to state the theorem, we need the

---

[4]The proofs involve technical arguments concerning roots of polynomials associated to the recurrence; in many cases one needs to assume all the recurrence coefficients are positive.

following definitions, as were presented in the above reference, but specialized to the m-gonal case. Let $\{a_n\}$ denote the m-gonal sequence with its unique decomposition as given in Definition 1.1. Let $I_n := [0, a_{mn+1})$ for all $n > 0$ and let $\delta(x - a)$ denote the Dirac delta functional, assigning a mass of 1 to $x = a$ and 0 otherwise.

• *Spacing gap measure:* We define the spacing gap measure of a $z \in I_n$ with $k(z)$ summands as

$$\nu_{z,n}(x) := \frac{1}{k(z) - 1} \sum_{j=2}^{k(z)} \delta(x - (\ell_j - \ell_{j-1})). \tag{1-12}$$

• *Average spacing gap measure:* Note that the total number of gaps for all $z \in I_n$ is

$$N_{\text{gaps}}(n) := \sum_{z=a_0}^{a_{mn+1}-1} (k(z) - 1). \tag{1-13}$$

The average spacing gap measure for all $z \in I_n$ is

$$\nu_n(x) := \frac{1}{N_{\text{gaps}}(n)} \sum_{z=a_0}^{a_{mn+1}-1} \sum_{j=2}^{k(z)} \delta(x - (\ell_j - \ell_{j-1}))$$

$$= \frac{1}{N_{\text{gaps}}(n)} \sum_{z=a_0}^{a_{mn+1}-1} (k(z) - 1) \nu_{z,n}(x). \tag{1-14}$$

Letting $P_n(g)$ denote the probability of a gap of length g among all gaps from the decompositions of all $z \in I_n$, we have

$$\nu_n(x) = \sum_{g=0}^{mn} P_n(g) \delta(x - g). \tag{1-15}$$

• *Limiting average spacing gap measure, limiting gap probabilities:* If the limits exist, we let

$$\nu(x) = \lim_{n \to \infty} \nu_n(x), \quad P(k) = \lim_{n \to \infty} P_n(k). \tag{1-16}$$

• *Indicator function for two gaps:* For $g_1, g_2 \geq 0$,

$$X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) := \#\big\{z \in I_n : a_{j_1}, a_{j_1+g_1}, a_{j_2}, a_{j_2+g_2} \text{ in } z\text{'s decomposition,}$$
$$\text{but not } a_{j_1+q}, a_{j_2+p} \text{ for } 0 < q < g_1, 0 < p < g_2\big\}. \tag{1-17}$$

• *Specific gap length probability:* Recall that $P_n(g)$ is the probability

$$P_n(g) := \frac{1}{N_{\text{gaps}}(n)} \sum_{i=1}^{mn+1-g} X_{i,i+g}(n). \tag{1-18}$$

Now can now state the result of the individual gap measure for the m-gonal case.

Theorem 1.8 (individual gap measure). *For $z \in I_n$, the individual gap measures $v_{z,n}(x)$ converge almost surely in distribution to the limiting gap measure $v(x)$.*

We give a proof of Theorem 1.8 in Section 5.

Longest gap. Another interesting problem is to determine the distribution of the longest gap between summands as $n \to \infty$. The structure of the legal m-gonal decompositions allows us to easily prove the following.

Theorem 1.9 (distribution of the longest gap). *Consider the m-gonal sequence $\{a_n\}$. Then as $n \to \infty$ the mean of the longest gap between summands in legal m-gonal decompositions of integers in $[a_n, a_{n+1})$ is $m \log_2(n/2m) + O_m(1)$, and the variance is $O_m(1)$.*

The proof of Theorem 1.9 is given in Section 6 and bypasses many of the technical arguments used in [Bower et al. 2015]. There the authors had to deduce properties of somewhat general associated polynomials; the nature of the legal m-gonal decompositions here allows us to immediately convert this problem to a simple generalization of the longest run of heads problem.

2. Recurrence relations and generating functions

Let $m \geq 1$. We can use the division algorithm to observe that the integer a_{mk+r} is the r-th integer in the bin b_{k+1} for $mk + r \geq 1$. Hence $1 \leq r \leq m$ denotes the location of the integer within its bin. We let the first bin b_0 contain the element $a_0 = 1$. Then for any $k \geq 0$, we let b_{k+1} denote the set of elements of the $(k+1)$-th bin. Namely

$$\underbrace{a_0}_{b_0}, \underbrace{a_1, a_2, \ldots, a_m}_{b_1}, \underbrace{a_{m+1}, a_{m+2}, \ldots, a_{2m}}_{b_2}, \ldots, \underbrace{a_{mk+1}, a_{mk+2}, \ldots, a_{m(k+1)}}_{b_{k+1}}, \ldots \qquad (2\text{-}1)$$

We can now begin our work in describing the terms of this sequence.

The following result, which follows immediately from the definition, is used in many of the proofs in this section. We record it here for easy reference.

Definition 2.1. Let Ω_n denote the largest integer with summands from each bin $b_0, b_1, b_2, \ldots, b_n$. Then

$$\Omega_n = \sum_{i=0}^{n} a_{mi}. \qquad (2\text{-}2)$$

The first result that makes use of Definition 2.1 is given below.

Lemma 2.2. *Let $m \geq 1$ and $k \geq 1$. If a_{mk+1} is the first entry in bin b_{k+1}, then $a_{mk+1} = a_{mk} + a_{m(k-1)+1}$.*

Proof. We note that since a_{mk+1} and $a_{m(k-1)+1}$ are the first numbers in the bins b_{k+1} and b_k, respectively, by (2-2) we have

$$a_{mk+1} = 1 + \Omega_k = 1 + \sum_{i=0}^{k} a_{mi}, \qquad (2\text{-}3)$$

$$a_{m(k-1)+1} = 1 + \Omega_{k-1} = 1 + \sum_{i=0}^{k-1} a_{mi}. \qquad (2\text{-}4)$$

Then (2-3) and (2-4) yield

$$a_{mk+1} = 1 + \Omega_k = 1 + a_{mk} + \Omega_{k-1} = a_{mk} + (1 + \Omega_{k-1}) = a_{mk} + a_{m(k-1)+1}, \quad (2\text{-}5)$$

as claimed. $\qquad \square$

We now prove a more general result.

Lemma 2.3. *If $k \geq 0$ and $1 \leq r \leq m$, then $a_{mk+r} = r a_{mk+1}$.*

Proof. First consider the bin b_1. As $b_0 = [a_0] = 1$, by construction of the m-gonal sequence it is straightforward to determine that $b_1 = [2, 4, \ldots, 2m]$ and $a_r = r a_1$ for all $1 \leq r \leq m$.

We proceed for bins b_k with $k \geq 1$ by induction on r, where $1 \leq r \leq m$. The base case $r = 1$ clearly holds.

Let $1 \leq x \leq m - 1$ and assume that for any $1 \leq r \leq x$, we have $a_{mk+r} = r a_{mk+1}$. We want to show that $a_{mk+x+1} = (x + 1) a_{mk+1}$. Recall that a_{mk+x+1} is the entry in bin b_{k+1} after a_{mk+x} and by definition a_{mk+x+1} is one more than the largest integer we can create using the elements of bins $b_0, b_1, \ldots, b_k$ along with the element a_{mk+x}. Using (2-2), we have

$$a_{mk+x+1} = 1 + a_{mk+x} + \Omega_k = 1 + a_{mk+x} + \sum_{i=0}^{k} a_{mi}. \qquad (2\text{-}6)$$

Since $1 + \Omega_k = a_{mk+1}$ and by the use of the induction hypothesis, (2-6) yields

$$a_{mk+x+1} = a_{mk+x} + a_{mk+1} = x a_{mk+1} + a_{mk+1} = (x + 1) a_{mk+1} \qquad (2\text{-}7)$$

as desired. $\qquad \square$

We now provide a closed formula for the terms of the m-gonal sequence.

Proposition 2.4. *Let $m \geq 2$, $k \geq 0$, and $1 \leq r \leq m$. Then $a_{mk+r} = 2r(m+1)^k$. For $m = 1$, we have $a_i = 2^i$.*

Proof. For the case where $m = 1$, each of our bins have size 1 and a legal decomposition has distinct summands. Thus the rule for legal decomposition is precisely a description of writing the positive integers in binary.

We will proceed by induction on k, the subscript on the bin, and r, the location of a_{mk+r} within the bin considered. The base case $k = 0$ and $r = 1$ gives the expected result, $a_{m \cdot 0 + 1} = 2(1)(m+1)^0 = 2$. We now assume that for some $k \geq 0$ and some r, $1 \leq r \leq m$, we have

$$a_{mk+r} = 2r(m+1)^k. \tag{2-8}$$

We need to show that the following two equations hold:

$$a_{mk+r+1} = 2(r+1)(m+1)^k, \tag{2-9}$$

$$a_{m(k+1)+r} = 2r(m+1)^{k+1}. \tag{2-10}$$

Suppose $1 \leq r \leq m-1$. To show (2-9) holds it suffices to observe that by Lemma 2.3 and our induction hypothesis we have

$$a_{mk+r+1} = (r+1)a_{mk+1} = (r+1) \cdot 2(1)(m+1)^k = 2(r+1)(m+1)^k. \tag{2-11}$$

When $r = m$, we use Lemma 2.2 and our induction hypothesis to quickly deduce

$$a_{mk+m+1} = 2(m+1)^{k+1}. \tag{2-12}$$

By iterating (2-11) until $r = m - 1$, we find that (2-12) holds. Then (2-10) holds by Lemma 2.3. □

The final result gives the recurrence relation stated in Theorem 1.4.

Corollary 2.5. *If $n > m$, then $a_n = (m+1)a_{n-m}$.*

Proof. Let $n > m$ and write $n = mk + r$, where $k \geq 1$ and $1 \leq r \leq m$. By Proposition 2.4, $a_n = a_{mk+r} = 2r(m+1)^k$ and $a_{n-m} = a_{m(k-1)+r} = 2r(m+1)^{k-1}$. So it directly follows that $a_n = (m+1)a_{n-m}$. □

Counting integers with exactly k summands. Koloğlu, Kopp, Miller and Wang [Koloğlu et al. 2011] introduced a very useful combinatorial perspective to attack Zeckendorf decomposition problems by partitioning the integers $z \in [F_n, F_{n+1})$ into sets based on the number of summands in their Zeckendorf decompositions. We use a similar technique to prove that the distribution of the average number of summands in the m-gonal decomposition displays Gaussian behavior.

Let $p_{n,k}$ denote the number of integers in $I_n := [0, a_{mn+1})$ whose m-gonal decomposition contains exactly k summands, where $k \geq 0$. We begin our analysis with the following result.

Proposition 2.6. *If $n, k \geq 0$, then*

$$p_{n,k} = \begin{cases} 1 & \text{if } k = 0, \\ m^k \binom{n}{k} + m^{k-1} \binom{n}{k-1} & \text{if } 1 \leq k \leq n+1, \\ 0 & \text{if } k > n+1. \end{cases} \tag{2-13}$$

Proof. Let $n, k \geq 0$. Observe that the unique integer in the interval $I_n = [0, a_{mn+1})$ which has zero summands is zero itself. Thus $p_{n,0} = 1$. Now if k is larger than the number of available bins, it would be impossible to have k summands, as one can draw no more than one summand per bin. Therefore $p_{n,k} = 0$ whenever $k > n + 1$.

We now show that for $1 \leq k \leq n + 1$, we have $p_{n,k} = m^k \binom{n}{k} + m^{k-1} \binom{n}{k-1}$. There are two cases to consider:

- **Case 1:** One of the k summands is chosen from b_0.

- **Case 2:** None of the k summands are chosen from b_0.

Case 1: Since one of the k summands is coming from b_0 there are $k - 1$ available summands to take from the bins $b_1, \ldots, b_n$. The number of ways to select $k - 1$ bins from n bins is $\binom{n}{k-1}$. As each of the bins $b_1, \ldots, b_n$ has exactly m elements and $|b_0| = 1$, once the $k - 1$ bins are selected, the number of ways to select an element from these bins is m^{k-1}. Thus the number of $z \in I_n$ which have exactly k summands with one summand coming from bin b_0 is $m^{k-1} \binom{n}{k-1}$.

Case 2: We choose k summands from any bin but b_0. Using a similar argument to that in Case 1, we can see that the total number of ways to select these k summands is $m^k \binom{n}{k}$.

As the two cases are disjoint, we have shown that the total number of integers in the interval I_n with exactly k summands is

$$p_{n,k} = m^k \binom{n}{k} + m^{k-1} \binom{n}{k-1} \tag{2-14}$$

as desired. $\qquad\square$

We also provide a recursive formula for the value of $p_{n,k}$, as it is used in the proof of Proposition 2.8.

Proposition 2.7. *If $0 < k < n + 1$, then $p_{n,k} = m p_{n-1,k-1} + p_{n-1,k}$.*

We omit the proof of Proposition 2.7, as it is a straightforward application of the combinatorial identity $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. With the recursive formula at hand, we now create a generating function for $p_{n,k}$.

Proposition 2.8. *Let*

$$F(x, y) := \sum_{n,k \geq 0} p_{n,k} x^n y^k \tag{2-15}$$

be the generating function of the $p_{n,k}$ arising from m-gonal decompositions. Then

$$F(x, y) = \frac{1 + y}{1 - (my + 1)x}. \tag{2-16}$$

Proof. Noting that $p_{n,k} = 0$ if either $n < 0$ or $k < 0$, using explicit values of $p_{n,k}$ and the recurrence relation from Proposition 2.7, and after some straightforward algebra, we obtain

$$F(x, y) = mxyF(x, y) + xF(x, y) + 1 + y. \tag{2-17}$$

From this, (2-16) follows. □

3. Gaussian behavior

To motivate this section's main result, we point the reader to the following experimental observations. Taking samples of 200,000 integers from the intervals $[0, 2 \cdot 4^{600})$, $[0, 2 \cdot 5^{600})$, $[0, 2 \cdot 6^{600})$ and $[0, 2 \cdot 7^{600})$, in Figure 3 we provide a histogram for the distribution of the number of summands in the m-gonal decomposition of these integers, when $m = 3, 4, 5$ and 6, respectively. Moreover, Figure 3 provides the histograms and Gaussian curves (associated to the respective value of m and n; the interval is $[0, a_{mn+1})$ so $n = 600$ in all experiments). In Table 1 we give the values of the predicted means and variances (as computed using Proposition 3.2), as well as the sample means and variances, for each of the cases considered.

From these observations it is expected that, for any $m \geq 1$, the distribution of the number of summands in the m-gonal decompositions of integers in the interval I_n will display Gaussian behavior. This is in fact the statement of Theorem 1.6. We begin by proving a technical result and follow it with the formulas for the mean and variance, which make use of some properties associated with the generating function for the $p_{n,k}$.

Proposition 3.1. *If $g_n(y)$ denotes the coefficient of x^n in $F(x, y)$, then*

$$g_n(y) = (1 + y)(my + 1)^n. \tag{3-1}$$

Proof. Using the fact that $F(x, y) = (1 + y)/(1 - mxy - x)$ we have by geometric series that

$$F(x, y) = \sum_{n=0}^{\infty} (1 + y)(my + 1)^n x^n. \tag{3-2}$$

Thus the coefficient of x^n in $F(x, y)$ is $(1 + y)(my + 1)^n$. □

We can now use $g_n(y)$ to find the mean and variance for the number of summands for integers $z \in I_n$.

Proposition 3.2. *Let Y_n be the number of summands in the m-gonal decomposition of a randomly chosen integer in the interval I_n, where each integer has an equal probability of being chosen. Let μ_n and σ_n^2 denote the mean and variance of Y_n. Then*

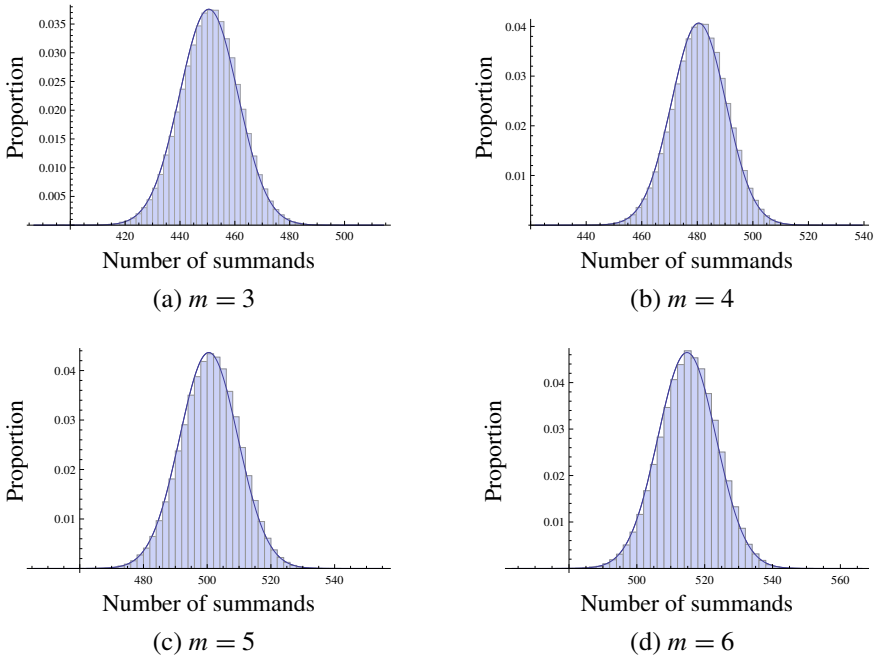$$\mu_n = \frac{nm}{m+1} + \frac{1}{2}, \quad \sigma_n^2 = \frac{nm}{(m+1)^2} + \frac{1}{4}. \tag{3-3}$$

(a) $m = 3$

(b) $m = 4$

(c) $m = 5$

(d) $m = 6$

Figure 3. Distributions for the number of summands in the m-gonal decomposition for a random sample with $n = 600$.

Figure	m	Predicted mean	Sample mean	Predicted variance	Sample variance
3(a)	3	450.50	450.49	112.75	112.34
3(b)	4	480.50	480.52	96.25	95.73
3(c)	5	500.50	450.49	83.58	83.38
3(d)	6	514.79	514.76	73.72	73.64

Table 1. Predicted means and variances versus sample means and variances for simulation from Figure 3.

Proof. By Propositions 4.7 and 4.8 in [Demontigny et al. 2014a] the mean and variance of Y_n are

$$\mu_n = \sum_{i=0}^{n} i \, P(Y_n = i) = \sum_{i=0}^{n} i \frac{p_{n,i}}{\sum_{k=0}^{n} p_{n,k}} = \frac{g_n'(1)}{g_n(1)}, \tag{3-4}$$

$$\sigma_n^2 = \sum_{i=0}^{n} (i - \mu_n)^2 P(Y_n = i) = \sum_{i=0}^{n} i^2 \frac{p_{n,i}}{\sum_{k=0}^{n} p_{n,k}} - \mu_n^2 = \frac{\frac{d}{dy}[y g_n'(y)]|_{y=1}}{g_n(1)} - \mu_n^2. \tag{3-5}$$

Our result follows from these formulas and the fact that $g_n(y) = (1 + y)(my + 1)^n$. $\square$

Normalize Y_n to $Y_n' = (Y_n - \mu_n)/\sigma_n$, where μ_n and σ_n are the mean and variance of Y_n respectively, as given in Proposition 3.2. We are now ready to prove that Y_n' converges in distribution to the standard normal distribution as $n \to \infty$.

Proof of Theorem 1.6. For convenience we set $r := t/\sigma_n$. Since

$$\sigma_n = \sqrt{\frac{nm}{(m+1)^2} + \frac{1}{4}},$$

we know that $r \to 0$ as $n \to \infty$ for any fixed value of t. Hence we will expand e^r using its power series expansion. We start with

$$M_{Y_n'}(t) = \frac{g_n(e^{t/\sigma_n})e^{-t\mu_n/\sigma_n}}{g_n(1)}. \tag{3-6}$$

Taking the logarithm of (3-6), we have

$$\log(M_{Y_n'}(t)) = \log[g_n(e^r)] - \log[g_n(1)] - \frac{t\mu_n}{\sigma_n}. \tag{3-7}$$

We proceed using Taylor expansions of the exponential and logarithmic functions to expand the following:

$$\begin{aligned}
\log[g_n(e^r)] &= \log(1+e^r) + n\log(me^r + 1) \\
&= \log\left(1 + \left(1 + r + \tfrac{1}{2}r^2\right)\right) + n\log\left(m\left(1 + r + \tfrac{1}{2}r^2\right) + 1\right) + O(r^3) \\
&= \log(2) + \tfrac{1}{2}\left(r + \tfrac{1}{2}r^2\right) - \tfrac{1}{8}\left(r + \tfrac{1}{2}r^2\right)^2 \\
&\quad + n\left(\log(m+1) + \frac{\left(mr + \tfrac{1}{2}mr^2\right)}{m+1} - \frac{\left(mr + \tfrac{1}{2}mr^2\right)^2}{2(m+1)^2}\right) + O(r^3) \\
&= \log(2(m+1)^n) + \tfrac{1}{2}r + \tfrac{1}{8}r^2 + \frac{nmr}{m+1} + \frac{nmr^2}{2(m+1)^2} + O(r^3). \tag{3-8}
\end{aligned}$$

From Proposition 3.1 we have $g_n(1) = 2(m+1)^n$, hence

$$\log[2(m+1)^n] = \log[g_n(1)]. \tag{3-9}$$

Substituting (3-8) and (3-9) and the values

$$\mu_n = \frac{nm}{m+1} + \frac{1}{2} \quad \text{and} \quad \sigma_n = \sqrt{\frac{nm}{(m+1)^2} + \frac{1}{4}}$$

into (3-7) yields

$$\log(M_{Y_n'}(t)) = \tfrac{1}{2}r + \tfrac{1}{8}r^2 + \frac{nmr}{m+1} + \frac{nmr^2}{2(m+1)^2} - \frac{t\left(\frac{nm}{m+1} + \frac{1}{2}\right)}{\sqrt{\frac{nm}{(m+1)^2} + \frac{1}{4}}} + O(r^3). \tag{3-10}$$

After some straightforward algebra we arrive at

$$\log(M_{Y'_n}(t)) = \tfrac{1}{2}t^2 + o(1); \tag{3-11}$$

the moment-generating proof of the central limit theorem now yields that the distribution converges to that of the standard normal distribution as $n \to \infty$. $\qquad\square$

4. Average gap measure

We now turn our attention to our final result in which we determine the behavior of gaps between summands. We begin with some preliminary notation in order to make our approach precise. For a positive integer $z \in I_n = [0, a_{mn+1})$ with m-gonal decomposition

$$z = a_{\ell_t} + a_{\ell_{t-1}} + \cdots + a_{\ell_1}, \tag{4-1}$$

where $\ell_1 < \ell_2 < \cdots < \ell_t$, we define the multiset of gaps of z as

$$\mathrm{Gaps}_n(z) := \{\ell_2 - \ell_1, \ell_3 - \ell_2, \ldots, \ell_t - \ell_{t-1}\}. \tag{4-2}$$

Our result will average over all $z \in [0, a_{mn+1})$ since we are interested in the average gap measure arising from m-gonal decompositions.

We follow the methods of [Beckwith et al. 2013; Bower et al. 2015]. In order to have a gap of length exactly g in the decomposition of z, there must be some index i such that a_i and a_{i+g} occur in z's decomposition, but a_j does not for any j between i and $i+g$. Thus for each i we count how many z have a_i and a_{i+g} but not a_j for $i < j < i+g$; summing this count over i gives the number of occurrences of a gap of length g among all the decompositions of z in our interval of interest. We want to compute the fraction of the gaps (of length g) arising from the decompositions of all $z \in I_n$. This probability is given by

$$P_n(g) := \frac{1}{(\mu_n - 1)a_{mn+1}} \sum_{z=0}^{a_{mn+1}-1} \sum_{i=0}^{mn+1-g} X_{i,g}(z), \tag{4-3}$$

where $X_{i,g}(z)$ is the indicator function.[5]

We are now ready to prove the result on the exponential decay in the distribution of gaps. The arguments in the proof of our main result (Theorem 1.7) are quite straightforward, but a bit tedious. To simplify our arguments, we write the gap length g as $m\alpha + \beta$, where $\alpha \geq 0$ and $0 \leq \beta < m$.

---

[5]For $1 \leq i, g \leq mn$ $X_{i,g}(z)$ denotes whether the decomposition of z has a gap of length g beginning at index i. That is, for $z = a_{\ell_t} + a_{\ell_{t-1}} + \cdots + a_{\ell_1}$,

$$X_{i,g}(z) = \begin{cases} 1 & \text{if } \exists \, j, 1 \leq j \leq t \text{ with } i = \ell_j \text{ and } i + g = \ell_{j+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof of Theorem 1.7. Let $g = m\alpha + \beta$, where $\alpha \geq 0$ and $0 \leq \beta < m$. We proceed by considering the following two cases:

- **Case 1:** $\alpha = 0$.
- **Case 2:** $\alpha > 0$.

Case 1: Let $\alpha = 0$. Hence $g = \beta$, where $0 < \beta < m$, and so our gap is less than the size of each bin b_i for $i > 0$ ($\alpha = 0$ and $\beta = 0$ would give us a gap of length 0 which is not m-gonal legal). We first consider gaps of length $g = \beta$ beginning at index 0. If $a_i = a_0$ then the only way to have a gap of length $g = \beta$ is if $a_{i+1} = a_\beta$. Now we are counting integers with m-gonal decompositions of the form $a_{\ell_t} + \cdots + a_{\ell_3} + a_\beta + a_0$. The number of $z \in I_n$ with summands $a_{\ell_3}, \ldots, a_{\ell_t}$ coming from bins $b_2, b_3, \ldots, b_n$ is $(m+1)^{n-1}$.

If $a_i = a_{mk+r}$, $k \geq 0$ and $1 \leq r \leq m$, then $a_{i+g} = a_{mk+r+\beta}$. Notice that the case $r + \beta \leq m$ cannot occur as this would force $\{a_i, a_{i+g}\} \subset b_{k+1}$, which leads to a decomposition that is not m-gonal legal, as only one summand can be taken per bin. Now if $r + \beta > m$, then $a_{i+g} \in b_{k+2}$.

Notice that in this case $a_i = a_{mk+r}$ is one of the largest β entries in bin b_{k+1} and thus there are β many choices for r, namely $r \in \{m-\beta+1, m-\beta+2, \ldots, m\}$.

Now we need to count the number of $z \in I_n \setminus \{0\} = [1, a_{mn+1})$ which have summands $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+2}$. We must have $0 \leq k \leq n - 2$.

As we have already used bins b_{k+1} and b_{k+2}, it follows from a straightforward combinatorial counting argument that the total number of integers $z \in I_n$ that can be created with summands $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+2}$ (with no summands in between) is given by

$$2\beta(m+1)^{n-2}, \tag{4-4}$$

where the factor of β comes from the β possible choices of a_i within the bin b_{k+1}. As we can vary k from 0 to $n - 2$, we find that the total number of integers $z \in I_n$ which contribute a gap between a_i ($i \neq 0$) and a_{i+g} (assuming that $r + \beta > m$) is given by

$$2(n-1)\beta(m+1)^{n-2}. \tag{4-5}$$

Observe that (4-5) does not account for the case when $i = 0$, which adds an extra factor of $(m+1)^{n-1}$. Therefore

$$\sum_{z=0}^{a_{mn+1}-1} \sum_{i=0}^{mn+1-g} X_{i,g}(z) = 2\beta(n-1)(m+1)^{n-2} + (m+1)^{n-1}. \tag{4-6}$$

Using Proposition 3.2 and Proposition 2.4 we have

$$(\mu_n - 1)a_{mn+1} = \left(\frac{mn}{m+1} - \frac{1}{2}\right)(2(m+1)^n) = (m+1)^{n-1}(2mn - (m+1)). \tag{4-7}$$

By (4-6) and (4-7), for $g = \beta$, with $0 \le \beta < m$, we have

$$P_n(g) = \frac{2\beta(n-1)}{(m+1)(2mn-m-1)} + \frac{1}{2mn-m-1}. \tag{4-8}$$

Now recall that $P(g) = \lim_{n \to \infty} P_n(g)$, so by letting $n \to \infty$ in (4-8) we have

$$P(g) = \frac{\beta}{m(m+1)} \tag{4-9}$$

whenever $g = \beta$ and $0 \le \beta < m$. This completes Case 1.

Case 2: Let $g = m\alpha + \beta$, where $\alpha \ge 1$ and $0 \le \beta < m$. First consider when $a_i = a_0$. If $\beta = 0$, then $a_{i+g} = a_{m\alpha} \in b_\alpha$. Otherwise, when $0 < \beta < m$, we have $a_{i+g} = a_{m\alpha+\beta} \in b_{\alpha+1}$. In the case of the former, the number of $z \in I_n$ with summands coming from bins $b_{\alpha+1}, b_{\alpha+2}, \dots, b_n$ is $(m+1)^{n-\alpha}$. In the latter case, the number of $z \in I_n$ with summands coming from bins $b_{\alpha+2}, b_{\alpha+3}, \dots, b_n$ is $(m+1)^{n-(\alpha+1)}$.

Now if $a_i = a_{mk+r}$, with $k \ge 0$ and $1 \le r \le m$, then $a_{i+g} = a_{m(k+\alpha)+r+\beta}$. Hence $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+1}$ whenever $1 \le r + \beta \le m$, or $a_{i+g} \in b_{k+\alpha+2}$ whenever $m < r + \beta < 2m$. Hence we consider the following subcases:

- **Subcase 1:** Let $1 \le r + \beta \le m$.

- **Subcase 2:** Let $m < r + \beta < 2m$.

Subcase 1: Let $1 \le r + \beta \le m$. In this case $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+1}$. Notice that in this case a_i must be one of the smallest $m - \beta$ entries in bin b_{k+1}. Namely $r = 1, 2, \dots, m - \beta$.

Now we need to count the number of $z \in I_n$ which have summands $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+1}$ (and no summands in between). For the decomposition to only have summands from bins $b_0, \dots, b_{k+1}, b_{k+\alpha+1}, \dots, b_n$, we must have $0 \le k \le n - (\alpha+1)$.

In order to have a gap created by $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+1}$, there must be no summands taken from b_j, where $k+1 < j < k+\alpha+1$. Again using a straightforward combinatorial counting argument, the total number of integers $z \in I_n$ which have summands $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+1}$ (with no summands in between) is given by

$$2(m-\beta)(m+1)^{n-\alpha-1}, \tag{4-10}$$

where the factor of $m - \beta$ comes from the $m - \beta$ possible choices of a_i within the bin b_{k+1}.

As we can vary k from 0 to $n - (\alpha+1)$ we find that the total number of integers $z \in I_n$ which contribute a gap between a_i ($i \ne 0$) and a_{i+g} in this case is

$$2(n-\alpha)(m-\beta)(m+1)^{n-\alpha-1}. \tag{4-11}$$

Subcase 2: Let $m < r + \beta < 2m$. In this case $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+2}$. Notice that in this case a_i can be any of the largest β entries in bin b_{k+1} so $a_i \in b_{k+1}$ and $a_{i+g} \in b_{k+\alpha+2}$. Namely $r = m+1-\beta, m+2-\beta, \dots, m$.

Using the same reasoning as in Subcase 1, we determine that the total number of integers meeting the conditions is

$$2\beta(n - \alpha - 1)(m + 1)^{n-\alpha-2}. \tag{4-12}$$

This completes Subcase 2.

We still need to account for the number of integers $z \in I_n$ which contribute a gap of length $g = m\alpha + \beta$ ($\alpha \geq 1$ and $0 \leq \beta < m$) beginning at index a_0. Recall we previously computed this quantity to be $(m + 1)^{n-\alpha-1}$ when $\beta > 0$ and the quantity is $(m + 1)^{n-\alpha}$ when $\beta = 0$.

Therefore we need to sum the values of (4-11) and (4-12) along with $(m+1)^{n-\alpha-1}$ when $\beta > 0$ to get

$$\sum_{z=0}^{a_{mn+1}-1} \sum_{i=0}^{mn+1-g} X_{i,g}(z) = 2(n-\alpha)(m-\beta)(m+1)^{n-\alpha-1}$$
$$+2\beta(n-\alpha-1)(m+1)^{n-\alpha-2}+(m+1)^{n-\alpha-1}. \tag{4-13}$$

By (4-13) and (4-7), for $g = m\alpha + \beta$, with $\alpha \geq 1$ and $0 < \beta < m$, we have

$$P_n(g) = \frac{2(n-\alpha)(m-\beta)}{(m+1)^\alpha(2mn-m-1)} + \frac{2\beta(n-\alpha-1)}{(m+1)^{\alpha+1}(2mn-m-1)}$$
$$+ \frac{1}{(m+1)^\alpha(2mn-m-1)}. \tag{4-14}$$

Now recall that $P(g) = \lim_{n\to\infty} P_n(g)$, so by letting $n \to \infty$ in (4-14) we have

$$P(g) = \frac{2(m-\beta)}{(m+1)^\alpha(2m)} + \frac{2\beta}{(m+1)^{\alpha+1}(2m)} = \frac{m+1-\beta}{(m+1)^{\alpha+1}} \tag{4-15}$$

whenever $g = m\alpha + \beta$, $\alpha \geq 1$ and $0 < \beta < m$.

Now for $\beta = 0$ we do not need to consider when $\alpha = 0$, as this would give us a gap $g = 0$. Also, as $1 \leq r \leq m$, we only meet the conditions of Subcase 1. Thus for $\alpha > 0$ and $\beta = 0$ we need to sum the values of (4-11) along with $(m + 1)^{n-\alpha}$ to get

$$\sum_{z=0}^{a_{mn+1}-1} \sum_{i=0}^{mn+1-g} X_{i,g}(z) = 2(n - \alpha)m(m + 1)^{n-\alpha-1} + (m + 1)^{n-\alpha}. \tag{4-16}$$

By (4-16) and (4-7), for $g = m\alpha$, with $\alpha \geq 1$, we have

$$P_n(g) = \frac{2(n-\alpha)m}{(m+1)^\alpha(2mn-m-1)} + \frac{1}{(m+1)^{\alpha-1}(2mn-m-1)}. \tag{4-17}$$

Now recall that $P(g) = \lim_{n\to\infty} P_n(g)$, so by letting $n \to \infty$ in (4-17) we have

$$P(g) = \frac{1}{(m+1)^\alpha} = \frac{m+1}{(m+1)^{\alpha+1}} \tag{4-18}$$

whenever $g = m\alpha + \beta$, $\alpha \geq 1$ and $\beta = 0$. □

5. Individual gap measure

In this section, we prove Theorem 1.8 by checking that the conditions given in [Dorward et al. 2015, Theorem 1.1] are satisfied in the m-gonal case. We restate this theorem below for ease of reference.

Theorem 5.1 [Dorward et al. 2015]. *For $z \in I_n$, the individual gap measures $\nu_{z,n}(x)$ converge almost surely in distribution to the average gap measure $\nu(x)$ if the following hold*:

(1) *The number of summands for decompositions of $z \in I_n$ converges to a Gaussian with mean $\mu_n = c_{\text{mean}} n + O(1)$ and variance $\sigma_n^2 = c_{\text{var}} n + O(1)$ for constants $c_{\text{mean}}, c_{\text{var}} > 0$, and $k(z) \ll n$ for all $z \in I_n$.*

(2) *We have the following, with $\lim_{n \to \infty} \sum_{g_1, g_2} \text{error}(n, g_1, g_2) = 0$:*

$$\frac{2}{|I_n| \mu_n^2} \sum_{j_1 < j_2} X_{j_1, j_1 + g_1, j_2, j_2 + g_2}(n) = P(g_1) P(g_2) + \text{error}(n, g_1, g_2). \qquad (5\text{-}1)$$

(3) *The limits in* (1-16) *exist.*

The above theorem is more general than we need, and in our particular case our interval of interest is $I_n = [0, a_{mn+1})$. Now observe that Proposition 3.2 and Theorem 1.6 ensure that the first criterion is met, and $k(z)$ is clearly at most $mn + 1$ and thus $k(z) \ll n$. In addition, the exponential decay seen in Theorem 1.7 shows that condition (3) is met. It remains to show that condition (2) of Theorem 5.1 holds.

Proposition 5.2. *We have*

$$\frac{2}{|I_n| \mu_n^2} \sum_{j_1 < j_2} X_{j_1, j_1 + g_1, j_2, j_2 + g_2}(n) = P(g_1) P(g_2) + \text{error}(n, g_1, g_2) \qquad (5\text{-}2)$$

and the sum of the error over all pairs (g_1, g_2) goes to zero as $n \to \infty$.

Proof. Let

$$g_1 = \alpha_1 m + \beta_1, \quad g_2 = \alpha_2 m + \beta_2, \quad j_1 = k_1 m + r_1, \quad j_2 = k_2 m + r_2,$$

where $0 \le \beta_1, \beta_2 < m$, $1 \le k_1, k_2 \le m$ and $k_1 < k_2$. Thus a_{j_1} and a_{j_2} are in bins b_{k_1+1} and b_{k_2+1} respectively. There are a number of cases to consider when determining $\sum_{j_1 < j_2} X_{j_1, j_1 + g_1, j_2, j_2 + g_2}(n)$, depending on whether or not $\alpha_1 = 0$ or $\alpha_2 = 0$. We include only the case when both $\alpha_1, \alpha_2 \ge 1$, as the other cases are similar.

There are several cases to consider. We will first consider the four cases which contribute to the main term and then bound the remaining cases. In all of the cases contributing to the main term we have $j_1 + g_1 \ne j_2$ and $j_1 \ne 0$ and thus we will suppose these conditions hold below.

Case 1: Let $1 \leq r_1 + \beta_1 \leq m$ and $1 \leq r_2 + \beta_2 \leq m$. First we determine the possible values of k_1 and k_2. In this case, the gap from g_1 spans $\alpha_1 + 1$ bins and the gap from g_2 spans $\alpha_2 + 1$ bins. Thus $0 \leq k_1 \leq n - \alpha_1 - \alpha_2 - 3$ and $k_1 + \alpha_1 + 2 \leq k_2 \leq n - \alpha_2 - 1$ and the number of choices for k_1 and k_2 is $\frac{1}{2}n^2 + O(n)$. Because of the restrictions of $1 \leq r_1 + \beta_1 \leq m$ and $1 \leq r_2 + \beta_2 \leq m$, within each bin there are $m - \beta_1$ choices for where to place a_{j_1} and $m - \beta_2$ choices for where to place a_{j_2}. Lastly, we determine the number of ways to choose the remaining elements for the decomposition. Because we are spanning $\alpha_1 + 1$ bins for the gap from g_1 and $\alpha_2 + 1$ bins for the gap from g_2, using a straightforward combinatorial counting argument, the number of integers that can be made using what remains is $2(m+1)^{n-\alpha_1-\alpha_2-2}$. Thus the total number of integers that can be made in this case is

$$2(m - \beta_1)(m - \beta_2)(m + 1)^{n-\alpha_1-\alpha_2-2}(n^2/2 + O(n))$$
$$= (m - \beta_1)(m - \beta_2)(m + 1)^{n-\alpha_1-\alpha_2-2}(n^2 + O(n)). \quad (5\text{-}3)$$

Through similar arguments we can obtain the remaining three cases that contribute to the main term.

Case 2: Let $1 \leq r_1 + \beta_1 \leq m$ and $m < r_2 + \beta_2 < 2m$. Then the number of integers that can be made in this case is

$$(m - \beta_1)\beta_2(m + 1)^{n-\alpha_1-\alpha_2-3}(n^2 + O(n)). \quad (5\text{-}4)$$

Case 3: Let $m < r_1 + \beta_1 < 2m$ and $1 \leq r_2 + \beta_2 \leq m$. Then the number of integers that can be made in this case is

$$(m - \beta_2)\beta_1(m + 1)^{n-\alpha_1-\alpha_2-3}(n^2 + O(n)). \quad (5\text{-}5)$$

Case 4: Let $m < r_1 + \beta_1 < 2m$ and $m < r_2 + \beta_2 < 2m$. Then the number of integers that can be made in this case is

$$\beta_1\beta_2(m + 1)^{n-\alpha_1-\alpha_2-4}(n^2 + O(n)). \quad (5\text{-}6)$$

The remaining cases occur when $j_1 + g_1 = j_2$ or $j_1 = 0$. In these cases, the number of choices for k_1 and k_2 is on the order of n instead of n^2 and thus the number of integers that can be made in these cases is $O(n(m + 1)^{n-\alpha_1-\alpha_2})$. Combining all cases, we have

$$\sum_{j_1 < j_2} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n) = n^2(m-\beta_1)(m-\beta_2)(m+1)^{n-\alpha_1-\alpha_2-2}$$
$$+ n^2(m-\beta_1)\beta_2(m+1)^{n-\alpha_1-\alpha_2-3}$$
$$+ n^2(m-\beta_2)\beta_1(m+1)^{n-\alpha_1-\alpha_2-3}$$
$$+ n^2\beta_1\beta_2(m+1)^{n-\alpha_1-\alpha_2-4}$$
$$+ O(n(m+1)^{n-\alpha_1-\alpha_2}). \quad (5\text{-}7)$$

Next, recall that by Proposition 3.2 we have $\mu_n = nm/(m+1) + \frac{1}{2}$. In addition, $|I_n| = a_{mn+1} = 2(m+1)^n$. Thus in our case we have

$$\frac{2}{|I_n|\mu_n^2} \sum_{j_1 < j_2} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n)$$

$$= \frac{1}{(m+1)^n \left(\frac{nm}{m+1} + \frac{1}{2}\right)^2} \sum_{j_1 < j_2} X_{j_1, j_1+g_1, j_2, j_2+g_2}(n)$$

$$= \left(\frac{1}{(m+1)^n \left(\frac{nm}{m+1} + \frac{1}{2}\right)^2}\right)\left(n^2(m - \beta_1)(m - \beta_2)(m+1)^{n-\alpha_1-\alpha_2-2}\right.$$

$$+ n^2(m - \beta_1)\beta_2(m+1)^{n-\alpha_1-\alpha_2-3} + n^2(m - \beta_2)\beta_1(m+1)^{n-\alpha_1-\alpha_2-3}$$

$$\left. + n^2\beta_1\beta_2(m+1)^{n-\alpha_1-\alpha_2-4} + O(n(m+1)^{n-\alpha_1-\alpha_2})\right)$$

$$= \left(\frac{1}{(m+1)^{\alpha_1+\alpha_2+2}m^2n^2 + O(n)}\right)\left(n^2(m - \beta_1)(m - \beta_2)(m+1)^2\right.$$

$$\left. + n^2(m - \beta_1)\beta_2(m+1) + n^2(m - \beta_2)\beta_1(m+1) + n^2\beta_1\beta_2\right)$$

$$+ O\left(\frac{1}{n(m+1)^{\alpha_1+\alpha_2}}\right). \tag{5-8}$$

Taking the limit as $n \to \infty$ and rearranging we obtain

$$\frac{m^2(m+1 - \beta_1)(m+1 - \beta_2)}{m^2(m+1)^{\alpha_1+\alpha_2+2}} = P(g_1)P(g_2). \tag{5-9}$$

The fact that the error term decays exponentially in g_1 and g_2 ensures that the error summed over all g_1 and g_2 goes to zero. $\qquad\square$

6. Longest gap

Using the techniques introduced by Bower, Insoft, Li, Miller and Tosteson [Bower et al. 2015], we can determine the mean and variance of the distribution of the longest gap between summands in the decomposition of integers in $[a_n, a_{n+1})$ as $n \to \infty$. There are no obstructions to using those methods; however, there are some bookkeeping issues due to the nature of our legal m-gonal decomposition. Specifically, we have to worry a little about the residue of the longest gap modulo m. This is a minor issue, as with probability 1 the longest gap is much larger than m and thus we will not have two items in the same bin. Rather than going through the technical argument, we instead give a very short proof that captures the correct main term of the mean of the longest gap, which grows linearly with $\log n$; our error is at the level of the constant term for the mean. We are able to handle the variance similarly, and compute that up to an error it is $O_m(1)$.

Proof of Theorem 1.9. The proof follows immediately from results on the longest run of heads in tosses of a fair coin; we sketch the details below. If a coin has probability p of heads and $q = 1 - p$ of tails, the expected longest run of heads and its variance in n tosses is

$$\mu_n = \log_{1/p}(nq) - \frac{\gamma}{\log p} - \frac{1}{2} + r_1(n) + \epsilon_1(n), \quad \sigma_n^2 = \frac{\pi^2}{6 \log^2 p} + \frac{1}{12} + r_2(n) + \epsilon_2(n);$$

(6-1)

here γ is Euler's constant, the $r_i(n)$ are at most .000016, and the $\epsilon_i(n)$ tend to zero as $n \to \infty$. Note the variance is bounded independently of n (by essentially 3.5); see [Schilling 1990] for a proof.

Note that for legal m-gonal decompositions we either have an element in a bin, or we do not. As all decompositions are equally likely, we see that these expansions are equivalent to flipping a coin with probability $\frac{1}{2}$ for each bin, and choosing exactly one of the m possible summands in that bin if we have a tail. As the probability that the longest gap is at the very beginning or very end of a sequence of coin tosses is negligible, we can ignore the fact that the first bin has size 1 and that we may only use part of the last bin if $n + 1$ is not a multiple of m. Thus gaps between bins used in a decomposition correspond to strings of consecutive heads.

As our integers lie in $[a_n, a_{n+1})$, we have $\lfloor n/m \rfloor + O(1) = n/m + O(1)$ bins (again, we ignore the presence or absence of the initial bin of length 1 or a partial bin at the end). We now invoke the results on the length of the longest run for tosses of a fair coin. For us, this translates not to a result on the longest gap between summands, but to a result on the longest number of *bins* between summands. It is trivial to pass from this to our desired result, as all we must do is multiply by m (the error will be at most $O(m)$ coming from the location of where the summands are in the two bins). This completes the proof of our claim on the mean; the variance follows similarly. □

Appendix: Proof of Theorem 1.5

Our proof is constructive. We build our sequence by only adjoining terms that ensure that we can *uniquely* decompose a number while never using more than one summand from the same bin. For a fixed $m \geq 1$ the sequence begins

$$\underbrace{1}_{b_0}, \underbrace{2, 4, 6, \ldots, 2m}_{b_1}, \underbrace{2(m+1), 4(m+1), 6(m+1), \ldots, 2m(m+1)}_{b_2}, \ldots. \quad \text{(A-1)}$$

Note we would not adjoin 7 because then there would be two legal m-gonal decompositions for 7, one using $7 = 7$ and the other being $7 = 6 + 1$. The next number in the sequence must be the smallest integer which cannot be legally decomposed using the current terms of the sequence.

We can now proceed with proof by induction. The integers $1, 2, 3, \ldots, 2m$ have unique decompositions as they are either in the sequence or are the sum of an even number from bin b_1 plus the 1 from bin b_0. The sequence continues

$$\ldots, \underbrace{a_{m(n-2)+1}, a_{m(n-2)+2}, \ldots, a_{m(n-1)}}_{b_{n-1}}, \underbrace{a_{m(n-1)+1}, \ldots, a_{mn}}_{b_n}, \underbrace{a_{mn+1}, \ldots, a_{m(n+1)}}_{b_{n+1}}, \ldots \qquad \text{(A-2)}$$

By induction we assume that there exists a unique decomposition for all integers $z \le a_{mn} + \Omega_{n-1}$, where Ω_{n-1} is the maximum integer that can be legally decomposed using terms in the set $\{a_0, a_1, a_2, a_3, \ldots, a_{m(n-1)}\}$.

By construction we have

$$\Omega_n = a_{mn} + \Omega_{n-1} = a_{mn} + a_{m(n-1)+1} - 1.$$

Let x be the maximum integer that can be legally decomposed using terms in the set $\{a_1, a_2, a_3, \ldots, a_{m(n-1)}\}$. Note $x = a_{m(n-1)+1} - 1$, as this is why we include $a_{m(n-1)+1}$ in the sequence.

Claim. *We have $a_{mn+1} = \Omega_n + 1$ and this decomposition is unique.*

By induction we know that Ω_n is the largest value that we can legally decompose using only terms in $\{a_0, a_1, a_2, \ldots, a_{mn}\}$. Hence we choose $\Omega_n + 1$ as a_{mn+1} and $\Omega_n + 1$ has a unique decomposition.

Claim. *Every $N \in [\Omega_n+1, \Omega_n+1+x] = [a_{mn+1}, a_{mn+1}+x]$ has a unique decomposition.*

We can legally and uniquely decompose the integers $1, 2, 3, \ldots, x$ using elements in the set $\{a_0, a_1, a_2, \ldots, a_{m(n-1)}\}$. Adding a_{mn+1} to the decomposition of any of these integers would still yield a legal m-gonal decomposition since a_{mn+1} is not in any of the bins $b_0, b_1, b_2, \ldots, b_{n-1}$. The uniqueness of these decompositions follows from the fact that if a_{mn+1} is not included as a summand, then the decomposition does not yield a number big enough to exceed $\Omega_n + 1$.

Claim. *We have $a_{mn+2} = \Omega_n + 1 + x + 1 = a_{mn+1} + x + 1$ and this decomposition is unique.*

By construction the largest integer that can be legally decomposed using terms $\{a_0, a_1, a_2, \ldots, a_{mn+1}\}$ is $\Omega_n + 1 + x$.

Claim. *Every $N \in [a_{mn+2}, a_{mn+2} + x]$ has a unique decomposition.*

First note that the decomposition exists, as we can legally and uniquely construct $a_{mn+2} + v$, where $0 \le v \le x$. For uniqueness, we note that if we do not use a_{mn+2}, then the summation would be too small.

Claim. *We know $a_{mn+2} + x$ is the largest integer that can be legally decomposed using terms $\{a_0, a_1, a_2, \ldots, a_{mn+2}\}$.*

This follows from construction. $\qquad \square$

Acknowledgments

References

[Alpert 2009] H. Alpert, "Differences of multiple Fibonacci numbers", *Integers* **9**:6 (2009), 745–749. MR Zbl

[Beckwith et al. 2013] O. Beckwith, A. Bower, L. Gaudet, R. Insoft, S. Li, S. J. Miller, and P. Tosteson, "The average gap distribution for generalized Zeckendorf decompositions", *Fibonacci Quart.* **51**:1 (2013), 13–27. MR Zbl

[Ben-Ari and Miller 2014] I. Ben-Ari and S. J. Miller, "A probabilistic approach to generalized Zeckendorf decompositions", preprint, 2014. arXiv

[Best et al. 2016] A. Best, P. Dynes, X. Edelsbrunner, B. McDonald, S. J. Miller, K. Tor, C. Turnage-Butterbaugh, and M. Weinstein, "Gaussian distribution of the number of summands in generalized Zeckendorf decomposition in small intervals", *Integers* **16** (2016), Paper A6. MR

[Bower et al. 2015] A. Bower, R. Insoft, S. Li, S. J. Miller, and P. Tosteson, "The distribution of gaps between summands in generalized Zeckendorf decompositions", *J. Combin. Theory Ser. A* **135** (2015), 130–160. MR Zbl

[Catral et al. 2014] M. Catral, P. L. Ford, P. E. Harris, S. J. Miller, and D. Nelson, "Generalizing Zeckendorf's theorem: the Kentucky sequence", *Fibonacci Quart.* **52**:5 (2014), 69–91.

[Catral et al. 2016a] M. Catral, P. L. Ford, P. E. Harris, S. J. Miller, and D. Nelson, "Legal decompositions arising from non-positive linear recurrences", preprint, 2016. arXiv

[Catral et al. 2016b] M. Catral, P. L. Ford, P. E. Harris, S. J. Miller, D. Nelson, Z. Pan, and H. Xu, "New behavior in legal decompositions arising from non-positive linear recurrences", preprint, 2016. arXiv

[Daykin 1960] D. E. Daykin, "Representation of natural numbers as sums of generalised Fibonacci numbers", *J. London Math. Soc.* **35** (1960), 143–160. MR Zbl

[Demontigny et al. 2014a] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, D. Moon, and U. Varma, "Generalizing Zeckendorf's theorem to f-decompositions", *J. Number Theory* **141** (2014), 136–158. MR Zbl

[Demontigny et al. 2014b] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, and U. Varma, "A generalization of Fibonacci far-difference representations and Gaussian behavior", *Fibonacci Quart.* **52**:3 (2014), 247–273. MR Zbl

[Dorward et al. 2015] R. Dorward, P. L. Ford, E. Fourakis, P. E. Harris, S. J. Miller, E. Palsson, and H. Paugh, "Individual gap measures from generalized Zeckendorf decompositions", preprint, 2015. arXiv

[Drmota and Gajdosik 1998] M. Drmota and J. Gajdosik, "The distribution of the sum-of-digits function", *J. Théor. Nombres Bordeaux* **10**:1 (1998), 17–32. MR Zbl

[Filipponi et al. 1994] P. Filipponi, P. J. Grabner, I. Nemes, A. Pethö, and R. F. Tichy, "Corrigendum to: "Generalized Zeckendorf expansions"", *Appl. Math. Lett.* **7**:6 (1994), 25–26. MR Zbl

[Grabner and Tichy 1990] P. J. Grabner and R. F. Tichy, "Contributions to digit expansions with respect to linear recurrences", *J. Number Theory* **36**:2 (1990), 160–169. MR Zbl

[Grabner et al. 1994] P. J. Grabner, R. F. Tichy, I. Nemes, and A. Pethö, "Generalized Zeckendorf expansions", *Appl. Math. Lett.* **7**:2 (1994), 25–28. MR Zbl

[Keller 1972] T. J. Keller, "Generalizations of Zeckendorf's theorem", *Fibonacci Quart.* **10**:1 (1972), 95–102, 111, 112. MR Zbl

[Koloğlu et al. 2011] M. Koloğlu, G. S. Kopp, S. J. Miller, and Y. Wang, "On the number of summands in Zeckendorf decompositions", *Fibonacci Quart.* **49**:2 (2011), 116–130. MR Zbl

[Lengyel 2006] T. Lengyel, "A counting based proof of the generalized Zeckendorf's theorem", *Fibonacci Quart.* **44**:4 (2006), 324–325. MR Zbl

[Miller and Wang 2012] S. J. Miller and Y. Wang, "From Fibonacci numbers to central limit type theorems", *J. Combin. Theory Ser. A* **119**:7 (2012), 1398–1413. MR Zbl

[Miller and Wang 2014] S. J. Miller and Y. Wang, "Gaussian behavior in generalized Zeckendorf decompositions", pp. 159–173 in *Combinatorial and additive number theory: CANT 2011 and 2012* (New York, NY, 2011–2012), edited by M. B. Nathanson, Springer Proceedings in Mathematics and Statistics **101**, Springer, New York, NY, 2014. MR Zbl

[Schilling 1990] M. F. Schilling, "The longest run of heads", *College Math. J.* **21**:3 (1990), 196–207. MR Zbl

[Steiner 2002] W. Steiner, "Parry expansions of polynomial sequences", *Integers* **2** (2002), Paper A14. MR Zbl

[Steiner 2005] W. Steiner, "The joint distribution of greedy and lazy Fibonacci expansions", *Fibonacci Quart.* **43**:1 (2005), 60–69. MR Zbl

[Zeckendorf 1972] E. Zeckendorf, "Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas", *Bull. Soc. Roy. Sci. Liège* **41** (1972), 179–182. MR Zbl

rdorward@oberlin.edu   *Department of Mathematics, Oberlin College, Oberlin, OH 44074, United States*

fordpl@bethanylb.edu   *Department of Mathematics and Physics, Bethany College, Lindsborg, KS 67456, United States*

erf1@williams.edu   *Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, United States*

pamela.e.harris@williams.edu *Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996, United States*

Current address: *Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, United States*

sjm1@williams.edu   Steven.Miller.MC.96@aya.yale.edu
                              Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, United States

palsson@vt.edu   *Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, United States*

Current address: *Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States*

hannah.paugh@usma.edu   *Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996, United States*

Loewner deformations driven by the Weierstrass function

Joan Lind and Jessica Robins

(Communicated by Michael Dorff)

The Loewner differential equation provides a way of encoding growing families of sets into continuous real-valued functions. Most famously, Schramm–Loewner evolution (SLE) consists of the growing random families of sets that are encoded via the Loewner equation by a multiple of Brownian motion. The purpose of this paper is to study the families of sets encoded by a multiple of the Weierstrass function, which is a deterministic analog of Brownian motion. We prove that there is a phase transition in this setting, just as there is in the SLE setting.

1. Introduction and results

Charles Loewner introduced his namesake differential equation in 1923, and the equation subsequently became an important and long-standing tool in complex analysis. Many decades later Oded Schramm rediscovered the Loewner equation as he was working on seemingly unrelated problems in probability and statistical physics. In 2000, Schramm introduced a family of random curves, which he called stochastic Loewner evolution, or SLE for short (and which have subsequently been renamed Schramm–Loewner evolution in Schramm's honor).

Roughly speaking, the Loewner equation provides a correspondence between 2-dimensional curves and continuous 1-dimensional functions (and a more careful description will be given in the next section). Schramm discovered that the SLE
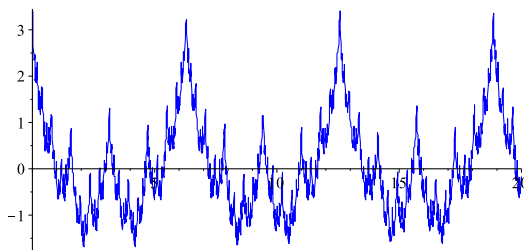
Figure 1. The Weierstrass function $W(t)$.

curves (the random 2-dimensional curves that he wanted to study) corresponded via the Loewner equation to a multiple of a well-known and much-loved random 1-dimensional function: Brownian motion. Thus, properties of Brownian motion could be leveraged to understand the SLE curves. Schramm's revolutionary work led not only to deep results in probability and theoretical physics, but it also inspired a renewed study of the Loewner equation. There has been particular interest in how geometric properties of the 2-dimensional curves may be encoded into the corresponding 1-dimensional functions.

SLE is often written SLE_{κ} to emphasize that it is an infinite family of random curves depending on a parameter $\kappa \geq 0$. In particular, under the Loewner correspondence, SLE_{κ} corresponds to the continuous function $\sqrt{\kappa} B(t)$, where $B(t)$ is Brownian motion. The SLE_{κ} curves come in three flavors, depending on the value of κ: when $\kappa \in [0, 4]$, then SLE_{κ} is a simple curve, when $\kappa \in (4, 8)$, then SLE_{κ} is a curve that hits back on itself, and when $\kappa \in [8, \infty)$, then SLE_{κ} is a space-filling curve [Rohde and Schramm 2005]. Thus there are three geometric phases, with sharp phase transitions at $\kappa = 4$ and $\kappa = 8$. See Figure 4, which illustrates the first two phases.

In this work we look at a deterministic analog of Brownian motion, the Weierstrass function, which, like Brownian motion, is continuous but nowhere-differentiable. In particular, we work with

$$W(t) = \sum_{n=0}^{\infty} 2^{-n/2} \cos(2^n t),$$

which is graphed in Figure 1. In comparison with SLE_{κ}, we seek to understand the 2-dimensional sets that correspond with a multiple of the Weierstrass function via the Loewner equation. We call this family of sets "the deformations driven by the Weierstrass function", and our main theorem establishes the existence of at least one phase transition, just as in the SLE setting. This transition from simple curve to nonsimple curve is illustrated in Figure 2, where we show approximations to the deformations driven by the Weierstrass function.
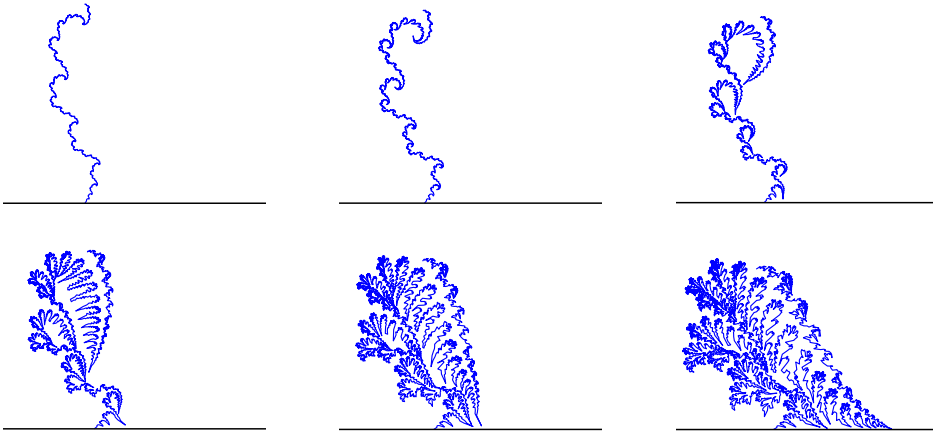
Figure 2. Simulations of the hulls generated by $cW(t)$ for $c = 0.8$ (top left), $c = 1$ (top middle), $c = 1.2$ (top right), $c = 1.4$ (bottom left), $c = 1.6$ (bottom middle), and $c = 1.8$ (bottom right).

Theorem 1.1. *The deformations driven by the Weierstrass function $W(t)$ exhibit a phase transition. In particular, when c is small enough, the hull generated by $cW(t)$ is a simple curve in $\mathbb{H} \cup \{cW(0)\}$, and this is not the case when c is large enough.*

In order to prove Theorem 1.1, we will need the following result, which gives a lower bound on the growth of the Weierstrass function near its local maxima.

Theorem 1.2. *Let $t_{m,k} = m\pi/2^k$ for $m, k \in \mathbb{N}$. If $0 < |h| \le 2^{-(k+7)}$, then*

$$W(t_{m,k}) - W(t_{m,k} + h) \ge 0.2\sqrt{|h|}.$$

This result implies that $W(t)$ has local maxima at the points $2^{-k}m\pi$. These times $2^{-k}m\pi$, which will feature in our proof of Theorem 1.1, correspond to the rightward-pointing "beaks" seen in the curves of Figure 2. One difference between Brownian motion and the Weierstrass function is that Brownian motion behaves similarly at its local maximums and local minimums, while the Weierstrass function favors its local maximums (that is, there is greater increase as one moves towards the local maximums than there is decrease moving towards the local minimums). This is also visually discernible in Figure 2 in the fact that there are obvious "beaks" to the right but not to the left.

Although we chose to focus on the Weierstrass function in this paper, we wish to note that our approach applies more generally. In fact, any $\mathrm{Lip}\left(\frac{1}{2}\right)$ function that has the behavior shown in Theorem 1.2 will exhibit a phase transition.

This paper is organized as follows. We discuss the Loewner equation in Section 2, with a focus on the particular aspects of the Loewner theory that will be needed

to prove Theorem 1.1. Section 3 regards the Weierstrass function and contains the proof of Theorem 1.2. In Section 4 we bring the Weierstrass function and the Loewner equation together to prove Theorem 1.1.

2. A look at the Loewner equation

In this section, we introduce the Loewner equation, consider some examples, and discuss the features of the Loewner equation that will be relevant for our work. We refer interested readers to the survey article [Gruzberg and Kadanoff 2004] and the references therein for more information about the Loewner equation and SLE.

Background and examples. The Loewner equation gives a correspondence between continuous, real-valued functions and certain growing families of sets in the complex plane. Given a function, we will describe how to obtain the family of sets via the Loewner equation. To that end, let λ be a continuous, real-valued function defined on $[0, T]$, and choose an initial point $z_0 \in \overline{\mathbb{H}} \setminus \{\lambda(0)\}$, where $\mathbb{H} = \{x + iy : y > 0\}$ denotes the upper half-plane. Then the chordal Loewner differential equation is the initial value problem

$$\frac{d}{dt}z(t) = \frac{2}{z(t) - \lambda(t)}, \quad z(0) = z_0. \tag{2-1}$$

A unique solution $z(t)$ exists on some time interval, by the existence and uniqueness theorem for differential equations. In fact, the solution $z(t)$ will continue to exist unless the denominator in (2-1) is zero, which occurs if $z(s) = \lambda(s)$ for some s. When this happens, we say that z_0 is captured by λ at time s. We define the hull at time t, notated K_t, to be the collection of captured points:

$$K_t = \{z_0 \in \overline{\mathbb{H}} : z(s) = \lambda(s) \text{ for some } s \le t\}.$$

This family of hulls, $\{K_t\}_{t \in [0,T]}$, is the increasing family of sets that correspond to $\lambda(t)$ via the Loewner equation. We call λ the driving function, and we say that K_t is generated by λ.

   We wish to take a moment to discuss the Loewner equation further in an informal manner. To begin, think of watching the movement of two particles in the plane. One particle moves only on $\mathbb{R}$ (and its position is given by $\lambda(t)$), and the other particle (described by $z(t)$) moves in $\overline{\mathbb{H}}$ but its movement is controlled by its relationship to the first particle via (2-1). To put a little action into our story, we think of the second particle as trying to escape from the first, while the first is trying to capture the second. To justify this storyline, let's suppose that $z_0 \in \mathbb{R}$, in which case both particles are moving along $\mathbb{R}$. Then (2-1) implies that the particle described by $z(t)$ is always moving away from the other particle (i.e., "trying to escape"). As we will
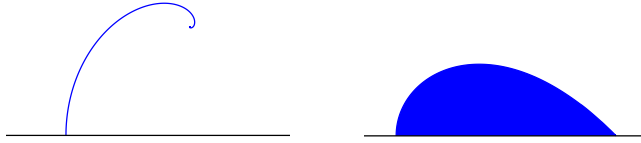
Figure 3. The hulls K_1 generated by $c - c\sqrt{1-t}$ for $c = 3$ (left) and $c = 5$ (right).

see later, if the particle described by $\lambda(t)$ moves quickly enough, it can catch up to the second particle and "capture" it (meaning that $\lambda(s) = z(s)$ at some time s).

We will briefly discuss some examples (and we refer the reader to [Kager et al. 2004] for the detailed analysis of these examples).

Example 1. When $\lambda(t) \equiv 0$, then $K_t = \{iy : 0 \leq y \leq 2\sqrt{t}\}$, a growing vertical line segment. To see why this might be true, we decompose (2-1) with $\lambda(t) \equiv 0$ into its real and imaginary parts:

$$\frac{d}{dt}\operatorname{Re}(z(t)) = \frac{2\operatorname{Re}(z(t))}{|z(t)|^2} \quad \text{and} \quad \frac{d}{dt}\operatorname{Im}(z(t)) = -\frac{2\operatorname{Im}(z(t))}{|z(t)|^2}.$$

This implies that $\operatorname{Im}(z(t))$ is decreasing, and $\operatorname{Re}(z(t))$ is increasing when $\operatorname{Re}(z_0) > 0$ and decreasing when $\operatorname{Re}(z_0) < 0$. In other words, points to the right of λ stay to the right, and points to the left of λ stay to the left. Thus the only possible points that could be captured by λ are those along the imaginary axis, since these points follow a downward trajectory toward λ.

Example 2. When $\lambda(t) = c\sqrt{t}$, then K_t is a growing line segment, beginning at 0. The angle between the line segment and $\mathbb{R}$ depends on c. This example is not as easy to justify as the first. One could either derive this result computationally (as done in [Kager et al. 2004]) or one could justify it using a scaling property of the Loewner equation. Neither approach, however, is relevant to the work in our paper, and we omit it.

In the first two examples, the hulls are growing simple curves in $\mathbb{H} \cup \{\lambda(0)\}$, by which we mean that there exists a simple curve $\gamma : [0, T] \to \mathbb{H} \cup \{\lambda(0)\}$ so that $K_t = \gamma([0, t])$. Initially, one might wonder if this is always true. The next example, however, shows us otherwise.

Example 3. Let $\lambda(t) = c - c\sqrt{1-t}$. For $0 < c < 4$, the hulls K_t are simple curves for all $t \in [0, 1]$. When $c \geq 4$, the same is true for the initial hulls; that is, for $t < 1$, K_t are simple curves. At $t = 1$, however, the geometry of the situation changes. Here the simple curve hits back on $\mathbb{R}$, and forms a "bubble", and so the final hull K_1 contains the curve, the points in $\mathbb{H}$ under the curve, and an interval from $\mathbb{R}$. See Figure 3.

Examples 2 and 3 each contain a family of examples, which we call a family of deformations. Our precise definition follows:
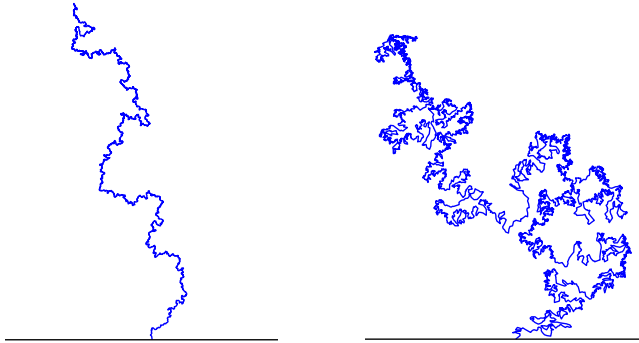
Figure 4. Simulations of an SLE_2 hull (the curve on the left) and an SLE_6 hull (the curve on the right together with all the bubbles formed).

Definition. Let λ be a continuous function defined on $[0, T]$. The family of deformations driven by λ is the family of hulls K_T^c generated by $c\lambda$ for $c > 0$.

Examples 2 and 3 gave the family of deformations driven by $\sqrt{t}$ and $1 - \sqrt{1-t}$. In Example 2, the hulls are simple curves for all values of c. However, there is a phase transition in Example 3: the hulls are simple curves for small c, but this fails to be the case for large c. Although this family is already well understood, we will prove the existence of this phase transition in Corollary 2.4 as an illustration of our method.

Example 4. SLE_κ, the best known example of a family of deformations, consists of the random hulls generated by $\sqrt{\kappa} B(t)$, where $B(t)$ is Brownian motion. As mentioned in the introduction, this family does exhibit phase transitions [Rohde and Schramm 2005]. In particular, when $\kappa \leq 4$, the hulls are simple curves (as illustrated in the left-hand picture of Figure 4), but this fails to be the case for $\kappa > 4$. When $4 < \kappa < 8$, the SLE_κ hull is the union of a random curve together with all the bubbles that are formed when the curve hits back on itself or on the real line (see the right-hand picture of Figure 4.) When $\kappa \geq 8$, there is a second phase transition, and the hulls become space-filling curves.

Criteria for hull behavior. In order to show that a family of deformations has a phase transition, we will need to be able to determine whether or not a hull is a simple curve, based on some feature of the driving function. In particular, our goal is to find two criteria; the first one (Theorem 2.1) will guarantee a simple-curve hull, and a second one (Proposition 2.2) will imply a nonsimple-curve hull. As an example, we will apply both of these to the driving function $c - c\sqrt{1-t}$ to verify the phase transition that we have discussed (see Corollary 2.4).

To formulate the first criterion, we need to define what it means for a function to be $\text{Lip}\left(\frac{1}{2}\right)$, also known as Hölder continuous with exponent $\frac{1}{2}$.
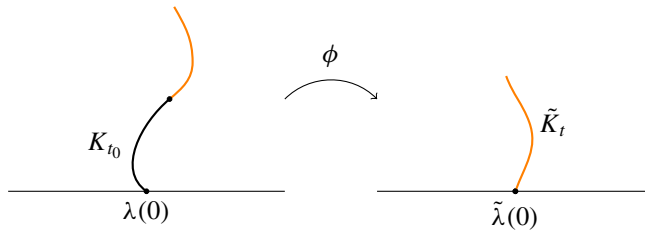
Figure 5. An illustration of the concatenation property.

Definition. A function $\lambda(t)$ defined on an interval $[0, T]$ is said to be a $\text{Lip}\left(\frac{1}{2}\right)$ function if there exists some $M < \infty$ so that

$$|\lambda(t) - \lambda(s)| \le M\sqrt{|t - s|}$$

for all $t, s \in [0, T]$. The smallest such M for which this holds is called the $\text{Lip}\left(\frac{1}{2}\right)$ norm of λ, notated $\|\lambda\|_{1/2}$.

Examples of $\text{Lip}\left(\frac{1}{2}\right)$ functions include $c\sqrt{t}$ and $c - c\sqrt{1 - t}$, both of which have $\text{Lip}\left(\frac{1}{2}\right)$ norm $|c|$. Further, any differentiable function will also be a $\text{Lip}\left(\frac{1}{2}\right)$ function. If $\lambda(t)$ is a $\text{Lip}\left(\frac{1}{2}\right)$ function with norm M, then $c\lambda(t)$ is also a $\text{Lip}\left(\frac{1}{2}\right)$ function and $\|c\lambda\|_{1/2} = |c|M$.

We use the following criterion when we want to guarantee we have a simple curve.

Theorem 2.1 [Lind 2005, Theorem 2]. *If λ is a $\text{Lip}\left(\frac{1}{2}\right)$ function with $\|\lambda\|_{1/2} < 4$, then the hulls generated by λ are all simple curves contained in $\mathbb{H} \cup \{\lambda(0)\}$.*

Next, we wish to formulate a criterion that will imply that a particular hull is not a simple curve. Consider Example 3, where the driving function is $c - c\sqrt{1 - t}$. If we compare the final hulls generated when $c = 3$ and when $c = 5$, we notice one key difference: the latter hull contains an interval along the real line, whereas the former contains no real-valued points except for the initial point. This means that in the second situation, there exists a real-valued point that is captured by λ at time 1. This observation, combined with the following property of the Loewner equation, leads to our second criterion, Proposition 2.2 below.

Concatenation property of the Loewner equation. Let λ be a continuous function defined on $[0, T]$ and let K_t be the hulls generated by λ. For $t_0 \in (0, T)$, let $\tilde{K}_t$ be the hulls generated by the time-shifted driving function $\tilde{\lambda}(t) = \lambda(t_0 + t)$ for $t \in [0, T - t_0]$. Then $\tilde{K}_t = \phi(K_{t_0+t} \setminus K_{t_0})$, where ϕ is the unique conformal map from $\mathbb{H} \setminus K_{t_0}$ onto $\mathbb{H}$ with the following normalization at infinity: $\phi(z) = z + O(1/z)$.

Note that a conformal map between two domains is a homeomorphism that is also complex differentiable. The concatenation property is illustrated in Figure 5. Here the black curve is K_{t_0}, and ϕ is a conformal map from $\mathbb{H} \setminus K_{t_0}$ (that is, $\mathbb{H}$

with the black curve removed) onto $\mathbb{H}$. The orange curve on the left represents $K_{t_0+t} \setminus K_{t_0}$, and the image of this set under ϕ is the orange curve on the right.

Proposition 2.2. *Let λ be a continuous function defined on $[0, T]$. Suppose there exists some $t_0 \in [0, T)$ and some $s \in (0, T - t_0]$ so that the time-shifted driving function $\tilde{\lambda}(t) = \lambda(t_0 + t)$ captures a real-valued point at time s. Then the hull K_{t_0+s} generated by λ is not a simple curve contained in $\mathbb{H} \cup \{\lambda(0)\}$.*

Proof. Since $\tilde{\lambda}(t)$ captures a real-valued point at time s, the corresponding hull $\tilde{K}_s$ must contain at least one point in $\mathbb{R}$ that is not the initial point $\tilde{\lambda}(0)$. But this implies that $\tilde{K}_s$ cannot be a simple curve in $\mathbb{H} \cup \{\tilde{\lambda}(0)\}$.

If K_{t_0+s} is a simple curve in $\mathbb{H} \cup \{\lambda(0)\}$, then $K_{t_0+s} \setminus K_{t_0}$ must be a simple curve in $\mathbb{H} \setminus K_{t_0}$. The concatenation property implies that $\tilde{K}_s$ is the image of $K_{t_0+s} \setminus K_{t_0}$ under a homeomorphism taking $\mathbb{H} \setminus K_{t_0}$ to $\mathbb{H}$, and so $\tilde{K}_t$ must also be a simple curve in $\mathbb{H} \cup \{\tilde{\lambda}(0)\}$. Since this is not the case, K_{t_0+s} cannot be a simple curve in $\mathbb{H} \cup \{\lambda(0)\}$. $\qquad\square$

As an example, we wish to apply Theorem 2.1 and Proposition 2.2 to the hulls generated by $c - c\sqrt{1-t}$ to prove that this family has a phase transition. The following lemma, which we will use again later, provides part of the argument.

Lemma 2.3. *Let $c \geq 4$, $\tau > 0$, and $a \in \mathbb{R}$, and set $b = \frac{1}{2}(-c + \sqrt{c^2 - 16})$. Then $x(t) = a + b\sqrt{\tau - t}$ is a solution to (2-1) when $\lambda(t) = a - c\sqrt{\tau - t}$. In particular, the driving function $a - c\sqrt{\tau - t}$ captures a real-valued point at time τ.*

Proof. To show the first statement, we must simply verify that

$$x'(t) = \frac{2}{x(t) - \lambda(t)}. \tag{2-2}$$

The left-hand side of (2-2) is $x'(t) = -b/(2\sqrt{\tau - t})$, and the right-hand side is

$$\frac{2}{(a + b\sqrt{\tau - t}) - (a - c\sqrt{\tau - t})} = \frac{2}{(b + c)\sqrt{\tau - t}}.$$

Thus (2-2) holds as long as $-b/2 = 2/(b + c)$, which can be easily verified.

The second statement follows from the fact that $x(\tau) = a = \lambda(\tau)$. $\qquad\square$

Now Theorem 2.1 and Proposition 2.2 imply the following:

Corollary 2.4. *The deformations driven by $1 - \sqrt{1-t}$ exhibit a phase transition. In particular, the hull K_1^c generated by $c - c\sqrt{1-t}$ is a simple curve in $\mathbb{H} \setminus \{0\}$ when $0 \leq c < 4$, and this is not the case when $c \geq 4$.*

Proof. Since $c - c\sqrt{1-t}$ is a $\mathrm{Lip}\left(\frac{1}{2}\right)$ function with norm c, Theorem 2.1 implies that K_1^c is a simple curve in $\mathbb{H} \setminus \{0\}$ when $0 \leq c < 4$.

Now suppose that $c \geq 4$. Then Lemma 2.3 implies that $c - c\sqrt{1-t}$ captures a real-valued point at time 1. Thus, applying Proposition 2.2 with $t_0 = 0$ and $s = 1$ gives that K_1^c is not a simple curve in $\mathbb{H} \setminus \{0\}$. $\qquad\square$

3. The Weierstrass function

Karl Weierstrass introduced the Weierstrass function in 1872,[1] giving the first published example of a continuous function that is nowhere differentiable. The function can be written as

$$F_{a,b}(t) = \sum_{n=0}^{\infty} a^n \cos(b^n t),$$

depending on two parameters $a \in (0, 1)$ and $b \geq 1/a$.[2] In this paper we will work with $b = 2$ and $a = 1/\sqrt{2}$, and so we define

$$W(t) = \sum_{n=0}^{\infty} 2^{-n/2} \cos(2^n t),$$

which is graphed in Figure 1. With this choice of parameters, G. H. Hardy [1916, Theorem 1.33] proved that $W(t)$ is a $\mathrm{Lip}\left(\frac{1}{2}\right)$ function. We will give a proof of Hardy's result that allows us to calculate the following upper bound on the $\mathrm{Lip}\left(\frac{1}{2}\right)$ norm of $W(t)$.

Proposition 3.1. *The* $\mathrm{Lip}\left(\frac{1}{2}\right)$ *norm of* $W(t) = \sum_{n=0}^{\infty} 2^{-n/2} \cos(2^n t)$ *satisfies*

$$\|W\|_{1/2} \leq 12.$$

This result complements Theorem 1.2, which gives a lower bound for a local version of the $\mathrm{Lip}\left(\frac{1}{2}\right)$ norm. The two results are illustrated in Figures 6 and 7, where we have plotted

$$\frac{W(t_{m,k}) - W(t_{m,k} + h)}{\sqrt{|h|}}$$

as a function of h for two choices of $t_{m,k} = 2^{-k}m\pi$. The left-hand picture for each figure shows $h \in [-1, 1]$ while the right-hand image is a "zoomed-in" picture with $h \in [-0.0001, 0.0001]$. We wish to point out a few features of these pictures. First, notice that the output values have an upper bound that is unaffected by the zooming. The existence of this global upper bound is a result of the bound on the $\mathrm{Lip}\left(\frac{1}{2}\right)$ norm in Proposition 3.1. A more interesting feature is the fact that the output values in the zoomed-in picture fall in a band that is bounded below. Theorem 1.2 guarantees that this lower bound exists for any $t_{m,k} = 2^{-k}m\pi$, once we zoom in far enough.

The bounds we obtain in Theorem 1.2 and Proposition 3.1 are far from optimal when compared with our experimental data. This is evident in the right-hand

---

[1] Weierstrass introduced his namesake function in a presentation on July 18, 1872, but his published work regarding this function [Weierstrass 1895] appeared later.

[2] These particular parameter values are due to Hardy [1916]; Weierstrass originally had more restrictions on the parameters.
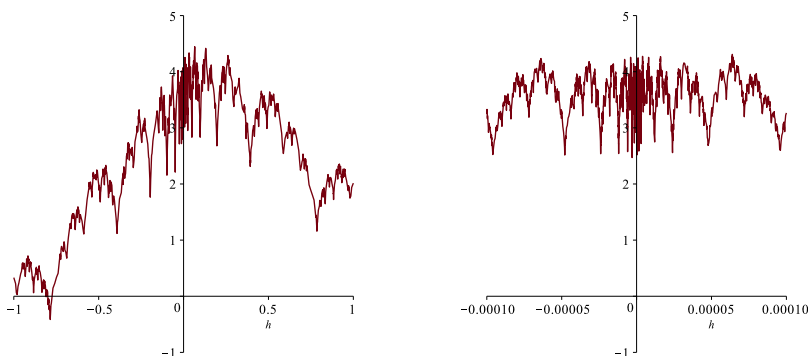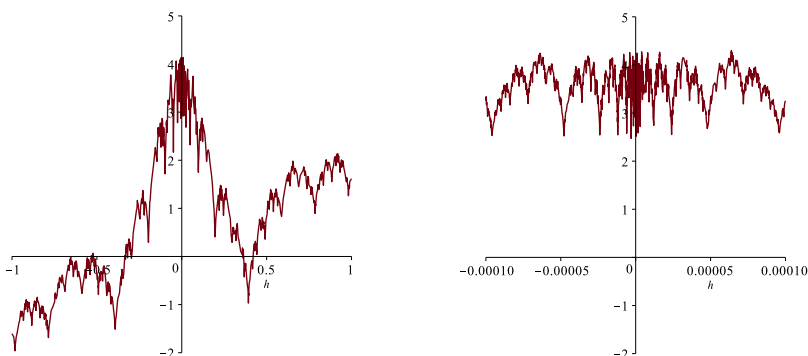
Figure 6. The graph of $(W(\pi/2) - W(\pi/2 + h))/\sqrt{|h|}$ for $h \in [-1, 1]$ (left) and for $h \in [-0.0001, 0.0001]$ (right).

pictures of Figures 6 and 7, which appear to be contained in a band between 2 and 5, a much more restrictive interval than the bounds we obtain of 0.2 and 12. The trade-off for our imprecision, however, is that our proofs are fairly straightforward.

Proof of Proposition 3.1. Set $a = 1/\sqrt{2}$. Note that

$$|W(t+h) - W(t)| \leq 2 \max_{s \in \mathbb{R}} |W(s)| = 2 \sum_{n=0}^{\infty} a^n = \frac{2}{1-a} \approx 6.8.$$

Therefore, when $|h| \geq 1$, we certainly have that $|W(t+h) - W(t)| \leq 12\sqrt{|h|}$.

For the rest of the proof, assume $0 < |h| < 1$. The trigonometric identity $\cos(x) - \cos(y) = -2\sin\left(\frac{1}{2}(x+y)\right)\sin\left(\frac{1}{2}(x-y)\right)$ implies that

$$|W(t+h) - W(t)| \leq 2 \sum_{n=0}^{\infty} a^n \left|\sin\left(\frac{1}{2}2^n(2t+h)\right)\right| \left|\sin\left(\frac{1}{2}2^n h\right)\right| \leq 2 \sum_{n=0}^{\infty} a^n |\sin(2^{n-1}h)|.$$

Figure 7. The graph of $(W(3\pi/8) - W(3\pi/8 + h))/\sqrt{|h|}$ for $h \in [-1, 1]$ (left) and for $h \in [-0.0001, 0.0001]$ (right).

We wish to find some integer p so that $2^{-p} \approx |h|$; then we will break the sum into two pieces based on p. The interval $(0, 1]$ can be decomposed into the union of the dyadic intervals $\left[\frac{1}{2}, 1\right], \left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{8}, \frac{1}{4}\right], \ldots$. Since $|h| \in (0, 1)$, we must have that $|h|$ is in one of these dyadic intervals; that is, there exists $p \in \mathbb{N}$ such that $2^{-p} \le |h| \le 2^{-p+1}$. Using this p, we split the sum into two pieces and bound each piece:

$$|W(t + h) - W(t)| \le 2 \sum_{n=0}^{p-1} a^n |\sin(2^{n-1}h)| + 2 \sum_{n=p}^{\infty} a^n |\sin(2^{n-1}h)|$$

$$\le 2 \sum_{n=0}^{p-1} a^n 2^{n-1}|h| + 2 \sum_{n=p}^{\infty} a^n$$

$$= |h| \frac{(2a)^p - 1}{2a - 1} + 2 \frac{a^p}{1 - a},$$

using the facts that $|\sin(x)| \le |x|$ and that

$$\sum_{n=0}^{p-1} r^n = \frac{r^p - 1}{r - 1}.$$

Since $2^{-p} \le |h| \le 2^{-p+1}$, we have that $2^p |h| \le 2$ and $a^p = \sqrt{2^{-p}} \le \sqrt{|h|}$. Thus

$$|W(t+h) - W(t)| \le \left(\frac{2}{2a-1} + \frac{2}{1-a}\right)a^p \le \left(\frac{2}{2a-1} + \frac{2}{1-a}\right)\sqrt{|h|} \approx 11.66\sqrt{|h|}. \quad \square$$

Proof of Theorem 1.2. Set $a = 1/\sqrt{2}$ and $t_{m,k} = m\pi/2^k$ for fixed $m, k \in \mathbb{N}$. Let $0 < |h| \le 2^{-(k+7)}$. As in the previous proof, we begin by applying the trigonometric identity $\cos(y) - \cos(x) = 2 \sin\left(\frac{1}{2}x + y\right) \sin\left(\frac{1}{2}x - y\right)$ to obtain

$$W(t_{m,k}) - W(t_{m,k} + h) = 2 \sum_{n=0}^{\infty} a^n \sin(2^n t_{m,k} + 2^{n-1}h) \sin(2^{n-1}h).$$

When $n \ge k + 1$, we know $2^n t_{m,k} = 2^{n-k}m\pi$ is a multiple of 2π, and so by the periodicity of the sine function, $\sin(2^n t_{m,k} + 2^{n-1}h) = \sin(2^{n-1}h)$. We split the sum into two pieces, the beginning and the tail:

$$W(t_{m,k}) - W(t_{m,k} + h) = B + T,$$

where

$$B = 2 \sum_{n=0}^{k} a^n \sin(2^n t_{m,k} + 2^{n-1}h) \sin(2^{n-1}h) \quad \text{and} \quad T = 2 \sum_{n=k+1}^{\infty} a^n \sin^2(2^{n-1}h).$$

We will have established the theorem once we show the two bounds

$$B \ge -0.31\sqrt{|h|} \quad \text{and} \quad T \ge 0.54\sqrt{|h|}. \tag{3-1}$$

We begin by showing the bound on B in (3-1), following similar reasoning to the previous proof:

$$B \geq -2 \sum_{n=0}^{k} a^n |\sin(2^{n-1}h)| \geq -2 \sum_{n=0}^{k} a^n 2^{n-1} |h| \geq -|h| \frac{(2a)^{k+1}}{2a-1},$$

since $|\sin(x)| \leq |x|$ and

$$\sum_{n=0}^{k} r^n = \frac{r^{k+1}-1}{r-1}.$$

Recall our assumption that $|h| \leq 2^{-(k+7)}$ and the fact that $2a = \sqrt{2}$. Therefore,

$$B \geq -\sqrt{|h|} \frac{\sqrt{|h|}\, 2^{(k+1)/2}}{\sqrt{2}-1} \geq -\sqrt{|h|} \frac{2^{-3}}{\sqrt{2}-1} \approx -0.302\sqrt{|h|}.$$

Now we will show the bound on T in (3-1). In proving this, we will assume, without loss of generality, that $h > 0$ (because $\sin^2(-x) = \sin^2(x)$). Since all the terms in T are positive, we can bound the infinite sum below by a partial sum; that is,

$$T = 2 \sum_{n=k+1}^{\infty} a^n \sin^2(2^{n-1}h) \geq 2 \sum_{n=k+1}^{p} a^n \sin^2(2^{n-1}h),$$

where $p \in \mathbb{N}$ satisfies $2^{-p} \leq h \leq 2^{-p+1}$. To show that this is well-defined, we need to know that $p \geq k + 1$. This follows from the assumption that $h \leq 2^{-(k+7)}$, which implies that $2^{-p} \leq 2^{-(k+7)}$ and subsequently $p \geq k + 7$.

Our next step is to bound the sine terms. When $0 \leq x \leq 1$, we have $\sin(x) \geq \sin(1) \cdot x$. To apply this to our situation, we need to verify that the argument of the sine terms is in the interval $[0, 1]$: for $n \leq p$ we have that $0 \leq 2^{n-1}h \leq 2^{p-1}h \leq 1$. Therefore

$$T \geq 2 \sum_{n=k+1}^{p} a^n \sin^2(2^{n-1}h) \geq 2 \sum_{n=k+1}^{p} a^n (\sin(1) \cdot 2^{n-1}h)^2 = \frac{\sin^2(1)}{2} h^2 \sum_{n=k+1}^{p} (4a)^n.$$

Set $r = 4a = 2^{3/2}$, and recall that

$$\sum_{n=k+1}^{p} r^n = \frac{r^{p+1} - r^{k+1}}{r-1}.$$

Since $h \geq 2^{-p}$,

$$h^2 = h^{3/2}\sqrt{h} \geq (2^{-p})^{3/2}\sqrt{h} = r^{-p}\sqrt{h}.$$

Putting this together, we have

$$T \geq \frac{\sin^2(1)}{2} h^2 \sum_{n=k+1}^{p} (4a)^n \geq \frac{\sin^2(1)}{2} r^{-p} \sqrt{h} \, \frac{r^{p+1} - r^{k+1}}{r-1}$$

$$= \sqrt{h} \, \frac{\sin^2(1)}{2} \frac{r - r^{k+1-p}}{r-1}$$

$$\geq \sqrt{h} \, \frac{\sin^2(1)}{2} \frac{r - r^{-6}}{r-1} \approx 0.547\sqrt{h},$$

where the final inequality follows from $p \geq k+7$. $\qquad\square$

4. Proof of the phase transition

We bring together the Loewner equation tools discussed in Section 2 and the properties of the Weierstrass function established in Section 3 to prove our main result.

Proof of Theorem 1.1. When $c < \frac{1}{3}$, Proposition 3.1 implies that $cW(t)$ is a $\mathrm{Lip}\big(\frac{1}{2}\big)$ function with norm below 4. Therefore, Theorem 2.1 ensures that the hull generated by $cW(t)$ is a simple curve in $\mathbb{H} \cup \{cW(0)\}$.

When $c \geq 20$, we will show that the hull generated by $cW(t)$ is not a simple curve in $\mathbb{H} \cup \{cW(0)\}$ by applying Proposition 2.2. To set the stage, let $c \geq 20$, let $k, m \in \mathbb{N}$, let $t_0 = 2^{-k} m\pi - 2^{-(k+7)}$, and define the time-shifted driving function $V(t) = cW(t_0 + t)$. Our proof will be complete once we prove that V captures a real-valued point at or before time $s = 2^{-(k+7)}$, and we will accomplish this by comparing V to a driving function that we understand well.

Let $t \in [0, s]$ and set $h = s - t$. Then by Theorem 1.2,

$$V(s) - V(t) = c\big(W(2^{-k}m\pi) - W(2^{-k}m\pi - h)\big) \geq c \cdot 0.2 \sqrt{h} \geq 4\sqrt{s-t}.$$

This implies that $V(t) \leq \lambda(t)$ for $\lambda(t) = V(s) - 4\sqrt{s-t}$. Notice also that $V(s) = \lambda(s)$. In other words, V and λ end at the same point, but V is below λ prior to this. Intuitively, this tells us that V must be moving quickly as $t \to s$, more quickly in fact than the function λ, which we know to capture a real-valued point (by Lemma 2.3), and so we should expect V will also capture a real-valued point. We simply need to adapt this intuition into a proof. We begin by appealing to Lemma 2.3, which implies that $x(t) = V(s) - 2\sqrt{s-t}$ is a solution (2-1) with driving function λ. Now let $u(t)$ be the solution to (2-1) with driving function V and initial condition $u(0) = x(0)$. We will assume that u is defined on $[0, s]$, because if not, that means that V has captured $u(0)$ before time s and we have nothing left to show.

Assume that $\tau \in [0, s]$ is a time so that $u(\tau) = x(\tau)$. Then since $V(\tau) \leq \lambda(\tau)$,

$$u'(\tau) = \frac{2}{u(\tau) - V(\tau)} \leq \frac{2}{x(\tau) - \lambda(\tau)} = x'(\tau).$$
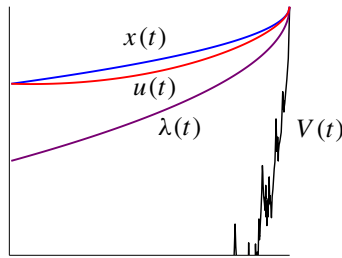
Figure 8. A sketch of the functions $x(t)$, $u(t)$, $\lambda(t)$ and $V(t)$ from the proof of Theorem 1.1.

So at any time when $u(\tau) = x(\tau)$, we have $x(t)$ is increasing more quickly than $u(t)$. This means that $u(t)$ can never pass $x(t)$, and so $u(t) \leq x(t)$ for all $t \in [0, s]$. Note that $u(0) = x(0) = V(s) - 2\sqrt{s} > V(s) - 4\sqrt{s} \geq V(0)$. In other words, $u(t)$ begins to the right of $V(t)$, and so $u(t)$ must remain to the right of $V(t)$ for as long as it is defined. Thus for all $t \in [0, s]$,

$$V(t) \leq u(t) \leq x(t),$$

as illustrated in Figure 8. At time s, we must have $V(s) \leq u(s) \leq x(s) = V(s)$. This implies $V(s) = u(s)$, meaning V has captured the real-valued point $u(0)$ at time s. $\square$

Acknowledgement

References

[Gruzberg and Kadanoff 2004] I. A. Gruzberg and L. P. Kadanoff, "The Loewner equation: maps and shapes", *J. Statist. Phys.* **114**:5–6 (2004), 1183–1198. MR Zbl

[Hardy 1916] G. H. Hardy, "Weierstrass's non-differentiable function", *Trans. Amer. Math. Soc.* **17**:3 (1916), 301–325. MR JFM

[Kager et al. 2004] W. Kager, B. Nienhuis, and L. P. Kadanoff, "Exact solutions for Loewner evolutions", *J. Statist. Phys.* **115**:3–4 (2004), 805–822. MR Zbl

[Lind 2005] J. R. Lind, "A sharp condition for the Loewner equation to generate slits", *Ann. Acad. Sci. Fenn. Math.* **30**:1 (2005), 143–158. MR Zbl

[Rohde and Schramm 2005] S. Rohde and O. Schramm, "Basic properties of SLE", *Ann. of Math.* (2) **161**:2 (2005), 883–924. MR Zbl

[Weierstrass 1895] K. Weierstrass, "Über continuirliche Functionen eines reellen Arguments, die für keinen Wert des letzeren einen bestimmten Differentialquotienten besitzen", pp. 71–74 in *Mathematische Werke von Karl Weierstrass*, Band 2, Mayer & Müller, Berlin, 1895. JFM

jlind@utk.edu   *Department of Mathematics, University of Tennessee, Knoxville, TN 37996, United States*

jlrobins15@gmail.com   *University of Tennessee, Knoxville, TN 37996, United States*

msp

Rank disequilibrium in multiple-criteria evaluation schemes

Jonathan K. Hodge, Faye Sprague-Williams and Jamie Woelk

(Communicated by Kenneth S. Berenhaut)

In multiple-criteria evaluation schemes, rank disequilibrium occurs when an evaluee is rated higher than other evaluees on some criteria and lower than other evaluees on other criteria. In this article, we investigate rank disequilibrium as it relates to the problem of aggregating scores on individual criteria into an overall evaluation. We adopt an axiomatic approach, defining the notion of a *rank aggregation function* and proposing a set of desirable properties — namely, independence, monotonicity, inclusivity, consistency, and equity — that rank aggregation functions may or may not satisfy. We prove that when there are more than three possible scores on each criterion, it is impossible to define a rank aggregation function that satisfies all of these properties. We then investigate potential resolutions to the problems posed by rank disequilibrium.

1. Introduction

According to Pruitt and Kim [2004, p. 24], *rank disequilibrium*, or *status inconsistency*, "exists when there are multiple criteria for assessing people's merit or contributions, and some people are higher on one criterion and lower on another criterion than others". Status inconsistency has been studied at length within the sociology and conflict resolution literature, particularly in regards to social stratification, intergroup conflict, and aggression (for example, [Engel 1988; Evan and Simmons 1969; Galtung 1964; Hernes 1969; Kriesberg 1998; Muller and Jukam 1983; Segal et al. 1970]). This article focuses on the phenomenon of rank disequilibrium within the specific context of multiple-criteria evaluation.

To illustrate, consider the common practice of evaluating employees by first assigning scores on a variety of criteria and then determining an overall performance rating by aggregating these scores in some way. For example, it is common within academia for faculty to be evaluated on teaching, research, and service. But what happens when Professor Smith is rated satisfactory in teaching, outstanding in research, and outstanding in service, while Professor Jones is rated outstanding in teaching, satisfactory in research, and satisfactory in service? Who should receive a higher overall rating? On the one hand, Professor Smith could argue that she is entitled to the higher rating, since she received higher marks on two of the three criteria. On the other hand, Professor Jones could argue that teaching is the most important criterion, and so she should receive the higher rating. The inconsistency between each professor's scores on the various criteria presents challenges to the evaluator who must aggregate the scores and determine an overall evaluation. Indeed, it is conceivable that, regardless of the final evaluations, one of the two professors will perceive that she has been treated inequitably.

Much of the prior research on rank disequilibrium has been empirical or philosophical in nature. In the present work, we take an axiomatic approach. In Section 2, we introduce the definitions that form the basis of our model. We define the notion of a *rank aggregation function*, which is our primary mechanism for aggregating marks on individual criteria into a single, overall evaluation. We then propose several desirable properties of rank aggregation functions — including *independence*, *monotonicity*, *inclusivity*, *consistency*, and *equity* — giving examples to illustrate each. Rank aggregation functions that satisfy all of these properties are said to be *ideal*.

In Sections 3 and 4, we consider conditions for the existence and uniqueness of ideal rank aggregation functions. We demonstrate the existence of a unique ideal rank aggregation function in the case where only three scores are possible for each criterion. We then prove that when more than three scores are allowed, no such function exists.

In Sections 5 and 6, we consider potential resolutions to the nonexistence (in most cases) of ideal rank aggregation functions. We show that suitable alternatives can be found if we are willing to sacrifice independence or accept a weaker form of equity.

In Section 7, we summarize our work and discuss questions for further research.

2. Definitions and examples

Our model assumes that a finite number of *evaluees* will receive one of a finite number of ratings, or *scores*, on each of a finite number of *criteria*.

We let C denote the set of evaluation criteria, where $|C| = n$. We let $Z = \{z_1, z_2, \ldots, z_m\}$ denote the set of possible scores for each criterion, where $m \geq 2$.

We assume that Z is totally ordered, with the ordering relation denoted by $\succeq$, where $z_m \succ z_{m-1} \succ \cdots \succ z_2 \succ z_1$. (Intuitively, z_m corresponds to the best score and z_1 corresponds to the worst score.)

A *score profile*, or *profile* for short, is an n-tuple of elements of Z — that is, an element of the Cartesian product

$$X = Z \times Z \times \cdots \times Z = Z^n.$$

This set X is called the *profile space*. For any profile $x \in X$, we let x_c denote the score from the c-th criterion (coordinate) of x. If, for some $x, y \in X$, we have $x_c \succeq y_c$ for all $c \in C$, then x is said to *dominate* y, denoted $x \gg y$. For any $x \in X$, we define $\min x$ and $\max x$ in a natural way — namely,

$$\min x = \min_{c \in C}\{x_c\} \quad \text{and} \quad \max x = \max_{c \in C}\{x_c\},$$

where the notions of minimum and maximum are defined with respect to the total order $\succeq$ on Z.

We use the notation $\vec{z}$ to denote the profile x for which $x_c = z$ for all $c \in C$. Such a profile is called a *uniform profile*. For every nonempty proper subset R of C, let $X_R = Z^{|R|}$ and $X_{-R} = Z^{n-|R|}$. For all $x \in X_R$ and $y \in X_{-R}$, we denote by (x, y) the profile that coincides with x for the criteria in R and coincides with y for the criteria not in R. In other words, $(x, y)_c = x_c$ for all $c \in R$ and $(x, y)_c = y_c$ for all $c \notin R$. To denote the restriction of a uniform profile $\vec{z}$ to the criteria in R, we write $\vec{z}_R$.

An *evaluee e* is an ordered pair $(x^e, \succeq_e)$, where $x^e \in X$ and $\succeq_e$ is a *monotone weak order* on X — that is, a weak order in which $x \succeq_e y$ whenever $x \gg y$. The first coordinate, x^e, represents the vector of scores assigned to e by the evaluation process (one for each evaluation criterion). The second coordinate, $\succeq_e$, represents e's perceived ordering of the possible profiles according to their relative value or desirability. (We will call such an ordering a *profile ordering*.) Monotonicity imposes a minimal assumption of rationality on each evaluee's profile ordering — for example, by prohibiting an evaluee from viewing a rating that is unsatisfactory in every category as more desirable than one that is satisfactory in every category. However if x and y are two profiles with the property that $x \not\gg y$ and $y \not\gg x$ (in which case $x_i \succ y_i$ and $y_j \succ x_j$ for some $i \neq j$), then the monotonicity condition imposes no restrictions on an evaluee's relative ordering of x and y. In this case, one can posit the existence of evaluees e_1 and e_2 for which $x \succ_{e_1} y$ and $y \succ_{e_2} x$.

The set of all possible evaluees is called the *evaluee space* and denoted $\mathcal{E}$. The set of all submultisets of $\mathcal{E}$ is denoted $\mathcal{P}(\mathcal{E})$.[1]

----

[1]Formally, our definitions must involve multisets to allow for the possibility of multiple evaluees having the same profile and profile ordering. However, from this point forward, we will, for convenience, use the simpler language of sets and subsets, keeping in mind that our results apply to (and sometimes require) multisets as well.

Definition 2.1. A *rank aggregation function* is a function that maps each pair (x, E), where x is a profile and E is a subset of $\mathcal{E}$, to an element of Z. In other words, a rank aggregation function is a mapping $f : X \times \mathcal{P}(\mathcal{E}) \to Z$.

For any rank aggregation function f, we will use the notation $f_E(x)$ to denote $f(x, E)$. This notation captures the idea that, for a *fixed* set of evaluees, we can view f as nothing more than a function on the profile space X. By definition, the function f is *anonymous*: it assigns the same overall score to any two evaluees who have the same profile. However, f may take into account information about the set of evaluees as a whole — including evaluees' profile orderings — so that if this information changes in any way, the scores assigned by f may also change. A rank aggregation function that assigns the same scores regardless of the set of evaluees is said to be *independent*. The next definition formalizes this idea and defines several additional desirable properties that a rank aggregation function might satisfy.

Definition 2.2. Let $f : X \times \mathcal{P}(\mathcal{E}) \to Z$ be a rank aggregation function. Then:

(i) f is *independent*, provided that $f_{E_1}(x) = f_{E_2}(x)$ for all $E_1, E_2 \in \mathcal{P}(\mathcal{E})$ and all $x \in X$.

(ii) f is *monotone*, provided that $x \gg y$ implies $f_E(x) \succeq f_E(y)$ for all $x, y \in X$ and all $E \in \mathcal{P}(\mathcal{E})$.

(iii) f is *inclusive*, provided that for each nonempty, proper subset R of C, there exists $E \in \mathcal{P}(\mathcal{E})$ and profiles $(x, u), (y, u)$, where $x, y \in X_R$ and $u \in X_{-R}$, such that $f_E(x, u) \neq f_E(y, u)$. (In this case, C is said to be *minimal* with respect to f.)

(iv) f is *consistent* with respect to $E \in \mathcal{P}(\mathcal{E})$, provided that $f_E(\vec{z}) = z$ for all $z \in Z$. If f is consistent with respect to *all* $E \in \mathcal{P}(\mathcal{E})$, then f is said to be *universally consistent*, or simply *consistent*.

(v) f is *equitable* with respect to $E \in \mathcal{P}(\mathcal{E})$, provided that $x \succ_e y$ implies $f_E(x) \succeq f_E(y)$ for all $e \in E$ and all $x, y \in X$. (If, for some evaluee $e \in E$, $x \succ_e y$ and $f_E(y) \succ f_E(x)$, then e is said to *perceive inequity*.) If f is equitable with respect to *all* $E \in \mathcal{P}(\mathcal{E})$, then f is said to be *universally equitable*, or simply *equitable*.

A rank aggregation function that is independent, monotone, inclusive, consistent, and equitable is said to be *ideal*. If f is an independent rank aggregation function, then we will denote by $f(x)$ the common value of $f_E(x)$ for all $E \in \mathcal{P}(\mathcal{E})$.

Example 2.3. Let f be the rank aggregation function defined by $f(x) = x_1$. (Note that f assigns to each profile the score received on the first criterion.)

By definition, f is independent and consistent. Furthermore, f is monotone since $x \gg y$ implies $x_1 \succeq y_1$.

Since f assigns scores based solely on the first criterion of each profile, it is intuitively obvious that f is not inclusive. To prove this formally, let R be any

nonempty, proper subset of C that does not include the first criterion. Then

$$f(x, u) = (x, u)_1 = u_1 = (y, u)_1 = f(y, u)$$

for all x, $y \in X_R$ and all $u \in X_{-R}$. So C is not minimal, and f is not inclusive.

Finally, we will show that f is not universally equitable. Let E be any set of evaluees containing an evaluee e for which

$$(z_1, z_m, \dots, z_m) \succ_e (z_m, z_1, \dots, z_1).$$

Such an evaluee exists since

$$(z_1, z_m, \dots, z_m) \gg (z_m, z_1, \dots, z_1) \quad \text{and} \quad (z_m, z_1, \dots, z_1) \gg (z_1, z_m, \dots, z_m).$$

Note, however, that since

$$f(z_m, z_1, \dots, z_1) = z_m \succ z_1 = f(z_1, z_m, \dots, z_m),$$

it follows that e perceives inequity, and f is not equitable with respect to E.

In conclusion, we have shown that f is independent, monotone, and consistent, but neither inclusive nor equitable. □

Example 2.4. Let f be the rank aggregation function defined by $f(x) = \max x$.

By definition, f is independent and consistent. We will show that f is inclusive and monotone, and f is equitable if and only if $m = 2$.

For inclusivity, let R be any proper, nonempty subset of C. Then

$$f(\vec{z}_1) = f((\vec{z}_1)_R, (\vec{z}_1)_{-R}) = z_1,$$

but

$$f((\vec{z}_m)_R, (\vec{z}_1)_{-R}) = z_m.$$

Since f assigns different scores to two profiles that differ only on R, and R was chosen arbitrarily, it follows that C is minimal with respect to f. Thus, f is inclusive.

For monotonicity, let x, $y \in X$ such that $x \gg y$. Then $x_c \succeq y_c$ for all $c \in C$, which implies

$$f(x) = \max x \succeq \max y = f(y).$$

Thus, f is monotone.

For equity, note that if $m \geq 3$, any evaluee e for which

$$\vec{z}_2 = (z_2, z_2, \dots, z_2) \succ_e (z_1, z_1, \dots, z_1, z_m)$$

will perceive inequity, since

$$f(z_1, z_1, \dots, z_1, z_m) = z_m \succ z_2 = f(\vec{z}_2).$$

(Intuitively, such an evaluee views a profile that receives the second worst score on each criterion as favorable to one that receives the worst score on all but one

of the criteria and the best score on the remaining criterion.) This example clearly breaks down if $m = 2$. Note that for an evaluee e to perceive inequity, it must be that, for some profiles x and y, we have $x \succ_e y$ and $f(y) \succ f(x)$. But when $m = 2$, $f(y) \succ f(x)$ only if $f(y) = z_2$ and $f(x) = z_1$. This, however, implies $x = \vec{z}_1$, a contradiction to the monotonicity of $\succ_e$ in light of the assumption that $x \succ_e y$.

To summarize, we have shown that f is independent, monotone, consistent, and inclusive. In the case that $m = 2$ (and only in this case), f is also equitable, and thus ideal.   □

Example 2.5. For a more concrete example, consider an evaluation process with three evaluation criteria ($n = 3$) and three possible scores ($m = 3$) for each criterion:

$$\text{outstanding} \succ \text{satisfactory} \succ \text{unsatisfactory}.$$

Define a rank aggregation function f as follows (abbreviating each score by its first letter):

$$f_E(x) = \begin{cases} O & \text{if } x \succ_e x^e \text{ for all } e \in E \text{ with } x^e \neq x, \\ U & \text{if } x^e \succ_e x \text{ for all } e \in E \text{ with } x^e \neq x, \\ S & \text{otherwise.} \end{cases}$$

Suppose Sally is being evaluated using f. Then the only way for Sally to be rated as outstanding is for every other evaluee whose profile is different from hers to view Sally's profile as the more favorable one. In essence, every evaluee must envy Sally's profile. Likewise, for Sally to be rated as unsatisfactory, every evaluee with a different profile than Sally's must view their own profile as favorable to hers.

Now consider the set E consisting of three evaluees — a, b, and c — for whom the following conditions hold:

- $x^a = (S, S, O)$ and $(S, S, O) \succ_a (S, O, S) \succ_a (O, S, S)$;
- $x^b = (S, O, S)$ and $(S, O, S) \succ_b (O, S, S) \succ_b (S, S, O)$;
- $x^c = (O, S, S)$ and $(O, S, S) \succ_c (S, S, O) \succ_c (S, O, S)$.

Note that, in this case, $f_E(S, S, O) = f_E(S, O, S) = f_E(O, S, S) = U$, since every evaluee favors their own profile over those of each of the other evaluees. (This is not a terribly surprising outcome, since evaluees tend to exhibit self-serving biases.)

Consider, however, another set E' of evaluees who, one might argue, are less tainted by their individual biases:

- $x^u = (S, S, O)$ and $(O, S, S) \succ_u (S, O, S) \succ_u (S, S, O)$;
- $x^v = (S, O, S)$ and $(O, S, S) \succ_v (S, O, S) \succ_v (S, S, O)$;
- $x^w = (O, S, S)$ and $(O, S, S) \succ_w (S, S, O) \succ_w (S, O, S)$.

It is easy to verify that $f_{E'}(S, S, O) = U$, $f_{E'}(S, O, S) = S$, and $f_{E'}(O, S, S) = O$. In other words, changing the set of evaluees changes the overall scores assigned by f to the profiles (S, O, S) and (O, S, S). Thus, f is *not* independent. This means that an evaluator who wished to use f would need to have some mechanism for ascertaining the evaluees' profile orderings. Moreover, as the example illustrates, misrepresentation of profile orderings — and even the influence of unintentional biases — could have a significant impact on the outcome of the evaluation. This phenomenon is analogous to the problem of insincere or strategic voting in social choice theory.

It is also worth noting that, while f may be equitable with respect to E (in this example, the profile orderings of a, b, and c are not specified completely enough to make a definitive conclusion), f is certainly *not* equitable with respect to E'; in particular, w perceives inequity since $(S, S, O) \succ_w (S, O, S)$ and

$$f_{E'}(S, O, S) = S \succ U = f_{E'}(S, S, O).$$

We leave it as an exercise to the reader to verify that f is monotone and inclusive, but not consistent. □

In the above examples, we considered rank aggregation functions that were monotone, but not necessarily equitable. In fact, because of the monotonicity requirement for profile orderings, the properties of monotonicity and equity (for rank aggregation functions) are closely related. This relationship can be made clear by examining the contrapositive of each property:

- *Monotonicity:* For all $x, y \in X$ and all $E \in \mathcal{P}(\mathcal{E})$, $f_E(y) \succ f_E(x)$ implies $x \gg\!\!\gg y$.
- *Equity:* For all $x, y \in X$, all $E \in \mathcal{P}(\mathcal{E})$, and all $e \in E$, $f_E(y) \succ f_E(x)$ implies $y \succeq_e x$.

The difference between monotonicity and equity therefore amounts to the difference between the conditions of $x \gg\!\!\gg y$ and $y \succeq_e x$. If $y \gg x$, then it is certainly the case that $y \succeq_e x$ for all e. But, as we have seen, $x \gg\!\!\gg y$ does not necessarily imply $y \gg x$. Likewise, it is also possible for $x \gg y$ and $y \succeq_e x$, which happens if and only if e is indifferent between x and y.

The arguments for monotonicity and equity are often similar, and so we will sometimes invoke the following lemma to prove both properties simultaneously:

Lemma 2.6. *Let f be a rank aggregation function, and suppose that for all $x, y \in X$, $f_E(y) \succ f_E(x)$ implies both (1) $x \gg\!\!\gg y$, and (2) $y \succeq_e x$ for every evaluee e. Then f is both monotone and (universally) equitable.*

Finally, it is important to recognize the difference between universality — particularly, as it applies to the properties of consistency and equity — and independence. Independence requires the scores assigned to profiles to depend only on the profiles themselves, and not on any properties of the underlying set of evaluees. However,

a rank aggregation function need not be independent to be universally consistent or equitable. The latter properties do not require the assigned scores to be invariant with respect to changes in the set of evaluees. Rather, these properties require only that the function be consistent/equitable with respect to *each* set of evaluees. Likewise, an independent rank aggregation function may be equitable with respect to some sets of evaluees, but not others. On the other hand, if an independent rank aggregation function is consistent with respect to any individual set of evaluees, then it will be universally consistent.

3. Existence of ideal rank aggregation functions

In the previous section, we saw an example of a rank aggregation function that was independent, monotone, consistent, and inclusive, but equitable (and thus ideal) only when $m = 2$ (Example 2.4). The rank aggregation function in Example 2.5 fared even worse: it was monotone and inclusive, but not independent, consistent, or equitable. This, of course, begs the question: What would an ideal rank aggregation function look like for $m \geq 3$? Does such a function even exist? The next theorem provides a partial answer to this question.

Theorem 3.1. *Let $m = 3$, and define the rank aggregation function f as*

$$f(x) = \begin{cases} z_1 & \text{if } x = \vec{z}_1, \\ z_3 & \text{if } x = \vec{z}_3, \\ z_2 & \text{otherwise.} \end{cases}$$

Then f is an ideal rank aggregation function.

Proof. By definition, f is independent and consistent. The argument for inclusivity is similar to that in Example 2.4.

For monotonicity and equity, we will use Lemma 2.6. So suppose $f(y) \succ f(x)$ for some $x, y \in X$. Then either $f(y) = z_3$ or $f(x) = z_1$. If the former, then $y = \vec{z}_3$ and $x \neq \vec{z}_3$. If the latter, then $x = \vec{z}_1$ and $y \neq \vec{z}_1$. In either case, $y \gg x$, which implies, by the monotonicity of profile orderings, that $y \succeq_e x$ for each evaluee e. Furthermore, $x \gg y$, and so f is both monotone and equitable. $\qquad\square$

Unfortunately, the function from Theorem 3.1 cannot be extended to cases where $m \geq 4$. In fact, for $m \geq 4$, the properties of consistency and equity are incompatible.

Theorem 3.2. *For $m \geq 4$, there does not exist a rank aggregation function that is consistent and equitable.*

Proof. Let $m \geq 4$, and assume to the contrary that f is a consistent and equitable rank aggregation function. Let $u = (z_1, z_m, z_m, \ldots, z_m)$, and let E contain evaluees e_1 and e_2 for which

$$\vec{z}_2 \succ_{e_1} u \succ_{e_2} \vec{z}_3.$$

Since f is equitable and consistent,

$$z_2 = f_E(\vec{z}_2) \succeq f_E(u) \succeq f_E(\vec{z}_3) = z_3.$$

This, however, is a contradiction to the fact that $z_3 \succ z_2$ by definition. □

Theorem 3.2 shows that, for $m \geq 4$, it is impossible for a consistent rank aggregation function to be equitable with respect to *every* possible set of evaluees. However, it does not rule out the possibility of finding a consistent rank aggregation function that is equitable with respect to *some* sets of evaluees. If one also desires independence, then this distinction is moot. (If f is independent, then since $f_E(\vec{z}_2) \succeq f_E(\vec{z}_3)$ for some set E of evaluees, the same ordering would hold for *every* set of evaluees, leading again to a contradiction.) If we are willing to sacrifice independence, then we may have more options. We will explore this possibility further in Section 5.

4. Cycles and uniqueness of ideal rank aggregation functions

In the proof of Theorem 3.2, two evaluees' profile orderings were combined in a way that forced a contradiction under the assumptions of both consistency and equity. This idea can be generalized as follows.

Definition 4.1. Let $E \in \mathcal{P}(\mathcal{E})$. Suppose there exist $e_1, \ldots, e_k \in E$ and $x^1, \ldots, x^k \in X$ such that

$$x^1 \succ_{e_1} x^2 \succ_{e_2} \cdots \succ_{e_{k-1}} x^k \succ_{e_k} x^1.$$

Then the sequence $x^1, \ldots, x^k, x^1$ is said to be a *strong k-cycle* with respect to E.

The following theorem is immediate:

Theorem 4.2. *Let $E \in \mathcal{P}(\mathcal{E})$, and let*

$$T_E = \{x \in X : x \text{ belongs to some strong cycle with respect to } E\}.$$

Let the relation $\sim_E$ on T_E be defined by $x \sim_E y$ if and only x and y belong to a common strong cycle. Then $\sim_E$ is an equivalence relation on T_E.

The next theorem can be viewed as a generalization of Theorem 3.2.

Theorem 4.3. *Let $E \in \mathcal{P}(\mathcal{E})$, and let f be a rank aggregation function that is equitable with respect to E. If $x \sim_E y$ for some x, $y \in X$, then $f_E(x) = f_E(y)$.*

Proof. Suppose f is equitable, and let $x \sim_E y$ for some x, $y \in X$. Then x and y belong to a common strong k-cycle with respect to E — that is,

$$x \succ_{e_1} \cdots \succ_{e_{j-1}} y \succ_{e_j} \cdots \succ_{e_k} x$$

for some $e_1, \ldots, e_k \in E$. But then equity requires that $f_E(x) \succeq f_E(y) \succeq f_E(x)$, which implies $f_E(x) = f_E(y)$. □

Two important corollaries follow.

Corollary 4.4. *Let $E \in \mathcal{P}(\mathcal{E})$. If $\vec{z}_i \sim_E \vec{z}_j$ for some $z_i, z_j \in Z$ with $i \neq j$, then there does not exist a rank aggregation function that is both consistent and equitable with respect to E.*

Proof. Suppose f is equitable with respect to E, and suppose also that $\vec{z}_i \sim_E \vec{z}_j$ for some $z_i, z_j \in Z$ with $i \neq j$. By Theorem 4.3, $f_E(\vec{z}_i) = f_E(\vec{z}_j)$. Since $i \neq j$, it is therefore impossible for f to be consistent with respect to E. $\square$

Corollary 4.5. *The function defined in Theorem 3.1 is the unique ideal rank aggregation function for $m = 3$.*

Proof. Let $m = 3$, and let f be an ideal rank aggregation function. Since f is consistent, $f(\vec{z}_1) = z_1$, $f(\vec{z}_2) = z_2$, and $f(\vec{z}_3) = z_3$. Choose $x \in X$ such that $x \neq \vec{z}_1, \vec{z}_2, \vec{z}_3$. We will show that there exists $E \in \mathcal{P}(\mathcal{E})$ such that $x \sim_E \vec{z}_2$, which will imply (by Theorem 4.3) that $f(x) = f(\vec{z}_2) = z_2$. Consider three cases.

<u>Case 1:</u> $x \gg \vec{z}_2$ and $\vec{z}_2 \gg x$. In this case, there exist evaluees e_1 and e_2 for which $\vec{z}_2 \succ_{e_1} x \succ_{e_2} \vec{z}_2$. Let E contain both e_1 and e_2. Then $x \sim_E \vec{z}_2$, as desired.

<u>Case 2:</u> $x \gg \vec{z}_2$. In this case, $x = ((\vec{z}_3)_R, (\vec{z}_2)_{-R})$ for some nonempty $R \subset C$. Let $y = ((\vec{z}_1)_R, (\vec{z}_3)_{-R})$. Note that (1) $\vec{z}_2 \gg y$ and $y \gg \vec{z}_2$, and (2) $x \gg y$ and $y \gg x$. By assumption, $x \gg \vec{z}_2$. Thus, there exist evaluees $e_1, e_2,$ and e_3 such that

$$\vec{z}_2 \succ_{e_1} y \succ_{e_2} x \succ_{e_3} \vec{z}_2.$$

It follows that $x \sim_E \vec{z}_2$ for any $E \in \mathcal{P}(\mathcal{E})$ containing $e_1, e_2,$ and e_3.

<u>Case 3:</u> $\vec{z}_2 \gg x$. In this case, $x = ((\vec{z}_1)_R, (\vec{z}_2)_{-R})$ for some nonempty $R \subset C$. Let $y = ((\vec{z}_3)_R, (\vec{z}_1)_{-R})$. By a similar argument as in Case 2, there exist evaluees $e_1, e_2,$ and e_3 such that

$$\vec{z}_2 \succ_{e_1} x \succ_{e_2} y \succ_{e_3} \vec{z}_2.$$

Thus, $x \sim_E \vec{z}_2$ for any $E \in \mathcal{P}(\mathcal{E})$ containing $e_1, e_2,$ and e_3.

In each case, there exists $E \in \mathcal{P}(\mathcal{E})$ such that $x \sim_E \vec{z}_2$; hence $f_E(x) = f_E(\vec{z}_2) = z_2$. Since f is independent, it follows that $f(x) = z_2$ for all $x \neq \vec{z}_1, \vec{z}_2, \vec{z}_3$. Thus, we have shown that f is identical to the function from Theorem 3.1. $\square$

5. Forfeiting independence

As we saw in Section 3, it is impossible to define a consistent rank aggregation function for $m \geq 4$ that is equitable with respect to *all* possible sets of evaluees. In fact, Corollary 4.4 tells us that consistency and equity are compatible only when the set E of evaluees is such that no two distinct uniform profiles belong to a common strong cycle. This necessary condition turns out to be sufficient as well, provided that we are willing to sacrifice independence.

In order to proceed, we first need the following definition:

Definition 5.1. Let $E \in \mathcal{P}(\mathcal{E})$, and let $x, y \in X$. Suppose there exist $e_1, \ldots, e_k \in E$ and $x^1, \ldots, x^{k-1} \in X$ such that

$$x \succ_{e_1} x^1 \succ_{e_2} \cdots \succ_{e_{k-1}} x^{k-1} \succ_{e_k} y.$$

Then x is said to *chain-dominate* y with respect to E, denoted $x \rightarrow_E y$.

Note that, for any $x, y \in X$, we have $x \sim_E y$ if and only if $x \rightarrow_E y$ and $y \rightarrow_E x$.

Theorem 5.2. *Let f be the rank aggregation function defined as follows:*

$$f_E(x) = \begin{cases} z_1 & \text{if } x = \vec{z}_1, \\ \vdots \\ z_i & \begin{array}{l} \text{if } f_E(x) \text{ is not defined above and either } \max x = z_i \\ \text{or there exists } w \in X \text{ such that } \max w = z_i \text{ and } w \rightarrow_E x, \end{array} \\ \vdots \\ z_m & \text{otherwise.} \end{cases}$$

Then f is monotone, inclusive, and equitable. Moreover, f is consistent with respect to any $E \in \mathcal{P}(\mathcal{E})$ for which $\vec{z}_i \sim_E \vec{z}_j$ only if $i = j$.

Proof. For monotonicity, suppose $x \gg y$ for some $x, y \in X$. We must show that $f_E(x) \succeq f_E(y)$ for all $E \in \mathcal{P}(\mathcal{E})$. Suppose $f_E(x) = z_i$. If $i = 1$, then $x = y = \vec{z}_1$, and so $f_E(x) = f_E(y) = z_1$. If $i = m$, then $f_E(x) = z_m \succeq f_E(y)$, as desired.

Now suppose $1 < i < m$. Then either (1) $\max x = z_i$, or (2) there exists $w \in X$ with $\max w = z_i$ such that $w \rightarrow_E x$. If $\max x = z_i$, then $\max y \preceq z_i$, and so $f_E(y) \preceq z_i = f_E(x)$, as desired. Suppose, on the other hand, that there exists $w \in X$ with $\max w = z_i$ such that $w \rightarrow_E x$. Since $w \rightarrow_E x$, there exist $e_1, \ldots, e_k \in E$ and $x^1, \ldots, x^{k-1} \in X$ such that

$$w \succ_{e_1} x^1 \succ_{e_2} \cdots \succ_{e_{k-1}} x^{k-1} \succ_{e_k} x.$$

But $x \gg y$, and so $x \succeq_{e_k} y$, which implies $x^{k-1} \succ_{e_k} y$. Thus $w \rightarrow_E y$, and so $f_E(y) \preceq z_i = f_E(x)$. In each case, $f_E(x) \succeq f_E(y)$. It follows that f is monotone.

For inclusivity, let R be any proper, nonempty subset of C. Then

$$f_E(\vec{z}_1) = f_E((\vec{z}_1)_R, (\vec{z}_1)_{-R}) = z_1,$$

but

$$f_E((\vec{z}_m)_R, (\vec{z}_1)_{-R}) \neq z_1.$$

Since f assigns different scores to two profiles that differ only on R, and R was chosen arbitrarily, it follows that C is minimal with respect to f. Thus, f is inclusive.

For equity, let $E \in \mathcal{P}(\mathcal{E})$, and suppose $x \succ_e y$ for some $x, y \in X$ and some $e \in E$. We must show that $f_E(x) \succeq f_E(y)$. Let $f_E(x) = z_i$. If $i = 1$, then $x = \vec{z}_1$, a contradiction to the fact that $x \succ_e y$. Therefore, $i > 1$. If $i = m$, then $f_E(x) = z_m \succeq f_E(y)$,

as desired. Now suppose $1 < i < m$. By the definition of f, either $\max x = z_i$ or there exists $w \in X$ with $\max w = z_i$ such that $w \to_E x$. If $\max x = z_i$, then the fact that $x \to_E y$ (since $x \succ_e y$) implies, by the definition of f, that $f_E(y) \preceq z_i$. If, on the other hand, there exists $w \in X$ with $\max w = z_i$ such that $w \to_E x$, then it must be that $w \to_E y$ as well (since $x \succ_e y$), and so $f_E(y) \preceq z_i$. In either case, $f_E(x) \succeq f_E(y)$. It follows that e cannot perceive inequity. Thus, f is equitable with respect to E.

For consistency, let $E \in \mathcal{P}(\mathcal{E})$ such that $\vec{z}_i \sim_E \vec{z}_j$ only if $i = j$. It suffices to show that, for each $1 < i \leq m$, there does not exist $w \in X$ such that $w \to_E \vec{z}_i$ and $\max w = z_j$ for some $j < i$. Suppose, to the contrary, that such a w exists. Since $w \to_E \vec{z}_i$, there exist $e_1, \ldots, e_k \in E$ and $x^1, \ldots, x^{k-1} \in X$ such that

$$w \succ_{e_1} x^1 \succ_{e_2} \cdots \succ_{e_{k-1}} x^{k-1} \succ_{e_k} \vec{z}_i.$$

Moreover, since $\max w = z_j$ and $j < i$, it follows that $\vec{z}_i \gg \vec{z}_j \gg w$. By the monotonicity of $\succeq_{e_1}$, we have $\vec{z}_i \succeq_{e_1} \vec{z}_j \succeq_{e_1} w$. Since $w \succ_{e_1} x^1$, this implies that $\vec{z}_i \succ_{e_1} x^1$ and $\vec{z}_j \succ_{e_1} x^1$. By the monotonicity of $\succeq_{e_k}$, we have $\vec{z}_i \succeq_{e_k} \vec{z}_j$, and so $x^{k-1} \succ_{e_k} \vec{z}_j$. It follows that

$$\vec{z}_i \succ_{e_1} x^1 \succ_{e_2} \cdots \succ_{e_{k-1}} x^{k-1} \succ_{e_k} \vec{z}_j \succ_{e_1} x^1 \succ_{e_2} \cdots \succ_{e_{k-1}} x^{k-1} \succ_{e_k} \vec{z}_i.$$

But then $\vec{z}_i \sim_E \vec{z}_j$ with $i \neq j$, a contradiction. Therefore, it must be the case that $f_E(\vec{z}_i) = z_i$ for all i, and f is consistent. □

6. Manifest inequity

In our investigations up to this point, we have not made a distinction between the *potential* for inequity and the actual *manifestation* of inequity in the scores assigned by a given rank aggregation function. The former involves a systemic or structural concern — namely, that a rank aggregation function *may* lead to ratings that are perceived by some to be inequitable, regardless of whether any specific evaluee receives one of the profiles involved in these potential inequities. However, perceived inequity involving profiles that are actually assigned to evaluees — what we will call *manifest* inequity — is especially problematic. In this section, we will consider the more modest goal of avoiding manifest inequity, defined formally below.

Definition 6.1. Let f be a rank aggregation function, and let $E \in \mathcal{P}(\mathcal{E})$. Suppose there exist $e_1, e_2 \in E$ such that $x^{e_1} \succ_{e_1} x^{e_2}$ and $f_E(x^{e_2}) \succ f_E(x^{e_1})$. Then

- x^{e_1} and x^{e_2} are called *manifest profiles* with respect to E; and

- e_1 is said to perceive *manifest inequity*.

A rank aggregation function for which no $e \in E$ perceives manifest inequity is said to be *weakly equitable* with respect to E.

Clearly, one way to avoid manifest inequity is to limit the profiles that are actually assigned to evaluees. In situations in which a single evaluator both assigns profiles to evaluees and chooses the rank aggregation function, this solution is both simple and practical. In fact, we will show that it is possible to define a rank aggregation function that is independent, monotone, consistent, inclusive, and weakly equitable with respect to any set of evaluees that satisfies the pairwise dominance condition defined below.

Definition 6.2. A set E of evaluees is said to be *pairwise dominant* if for all $e_1, e_2 \in E$, either $x^{e_1} \gg x^{e_2}$ or $x^{e_2} \gg x^{e_1}$. Likewise, a set S of profiles is said to be pairwise dominant if for all $x, y \in S$, either $x \gg y$ or $y \gg x$.

Theorem 6.3. *Let f be the rank aggregation function defined by $f(x) = \max x$. Then f is independent, consistent, inclusive, monotone, and weakly equitable with respect to any pairwise dominant set E of evaluees.*

Proof. By definition, f is independent and consistent. The proof of inclusivity and monotonicity is given in Example 2.4.

Now let E be any pairwise dominant set of evaluees. We must show that f is weakly equitable with respect to E. Suppose $x^{e_1} \succ_{e_1} x^{e_2}$ for some $e_1, e_2 \in E$. Since E is assumed to be pairwise dominant, either $x^{e_1} \gg x^{e_2}$ or $x^{e_2} \gg x^{e_1}$. If $x^{e_2} \gg x^{e_1}$, then the monotonicity of $\succeq_{e_1}$ implies $x^{e_2} \succeq_{e_1} x^{e_1}$, a contradiction. Therefore, $x^{e_1} \gg x^{e_2}$, and so

$$f(x^{e_1}) = \max x^{e_1} \geq \max x^{e_2} = f(x^{e_2}),$$

as desired. $\qquad\square$

Theorem 6.3 raises two interesting combinatorial questions. First, what is the cardinality of a maximal pairwise dominant set of profiles? Second, how many such sets exist? The next theorem provides answers to these questions.

Theorem 6.4. *Let S be a maximal pairwise dominant set of profiles — i.e., a set of profiles S that is pairwise dominant and is not a proper subset of any other pairwise dominant set. Then*

$$|S| = n(m - 1) + 1.$$

Moreover, the number of distinct pairwise dominant sets of profiles is equal to

$$\frac{(n(m - 1))!}{(m - 1)!^n}.$$

Proof. First, note that $(X, \gg)$ is a poset, and x covers y in $(X, \gg)$ if and only if there exists $c \in C$ such that (1) $x_i = y_i$ for all $i \neq c$, and (2) $x_c = z_k$ and $y_c = z_{k-1}$ for some $1 < k \leq m$. The cardinality of any maximal pairwise dominant set is the same as the length of a chain of maximum length in $(X, \gg)$. Such a chain must have maximum element $\vec{z}_m$, with each subsequent element formed by changing the

n	m	m^n	$n(m-1)+1$	$\dfrac{(n(m-1))!}{(m-1)!^n}$
3	3	27	7	90
3	4	64	10	1680
3	5	125	13	34650
4	3	81	9	2520
4	4	256	13	369600
4	5	625	17	63063000

Table 1. Combinatorial results for small m, n.

preceding element's score from z_k to z_{k-1} on exactly one criterion. Since there are n criteria, and the score on each criterion may be decreased (from z_k to z_{k-1}) exactly $m-1$ times before reaching z_1, it follows that a chain in $(X, \gg)$ can have length at most $n(m-1)+1$.

Each maximal chain (and, hence, each maximal pairwise dominant set) is uniquely determined by the order in which the criteria are decreased. Thus, each maximal pairwise dominant set corresponds to a sequence of $n(m-1)$ elements of C, where each of the n elements of C appears $m-1$ times. The standard formula for counting permutations with repetition thus implies that there are

$$\frac{(n(m-1))!}{(m-1)!^n}$$

such sequences, as desired.   □

Table 1 illustrates the results of Theorem 6.4 for some small values of m and n. While it is clear that pairwise dominant sets of profiles are small in comparison to the total number of profiles (m^n), it is also the case that there are many such sets to choose from. Therefore, an evaluator who wishes to take advantage of Theorem 6.3 must accept some fairly severe restrictions, but can satisfy these restrictions in a number of different ways.

7. Summary and conclusions

Rank disequilibrium can be a significant factor in social and intergroup conflict. In this article, we investigated rank disequilibrium in the context of multiple-criteria evaluation, using an axiomatic approach to show that, in general, it is impossible to define a function that aggregates scores on individual criteria while satisfying a relatively small set of desirable properties. In particular, we showed the notions of consistency and equity are generally incompatible. It seems perfectly reasonable to expect evaluees who receive the same score on every criterion to also receive that

score as an overall rating. However, this apparently innocuous requirement opens the door for evaluees to perceive inequity.

The problems illuminated by our analysis are not insoluble. When evaluators are limited to only three possible scores (say, outstanding, satisfactory, and unsatisfactory), an ideal rank aggregation function can be found. However, the only such function assigns a overall rating of satisfactory to almost all profiles, rating evaluees as outstanding or unsatisfactory only when they receive these respective scores on *every* criterion.

Another potential solution is to allow the evaluator to use information about evaluees' individual profile orderings in order to assign scores that minimize the potential for perceived inequity. An evaluator who implements such a system is likely to be motivated more by political considerations than an actual concern for equity. Indeed, doing so requires a great deal of effort (to ascertain reliable information about evaluees' relative orderings of the various profiles) without an absolute guarantee that the ideals of consistency and equity will be achieved.

A more practical solution involves simply limiting the profiles assigned to evaluees to ensure that, for any pair of profiles assigned, one profile is viewed as more desirable by *all* rational evaluees (that is, all evaluees whose profile orderings meet the modest condition of monotonicity.) An evaluator need not have foreknowledge of the evaluees' views in order to adopt this strategy, but must be willing to possibly rate evaluees insincerely in order to avoid assigning profiles that fall outside the allowed set. She must also settle for the weaker goal of avoiding *manifest* inequity, rather than *all* perceived inequity. Broader systemic or structural concerns may still persist, but the most glaring perceptions of inequity will be eliminated.

Our investigations rest on several assumptions that could be relaxed in future work. For example, we have required the same rating scale (i.e., the same set of scores) to be used for each criterion, as well as for the overall evaluation. We have also assumed the set of scores to be discrete. In other multiple-criteria decision contexts (such as voting on referenda; see [Bradley et al. 2005], for example), the distinction between discrete and continuous alternative sets has been shown to be significant.

Our model does not incorporate any assumptions about the relative importance of the evaluation criteria, nor does it address potential interdependencies among criteria. (In early critiques of research on rank disequilibrium, Doreian and Stockman [1969] and Hartman [1974] raise related concerns.) Given the importance of separability in economics and social choice theory [Bradley et al. 2005; Gorman 1968; Hodge and Schwallier 2006], these areas would seem to warrant further investigation.

References

[Bradley et al. 2005] W. J. Bradley, J. K. Hodge, and D. M. Kilgour, "Separable discrete preferences", *Math. Social Sci.* **49**:3 (2005), 335–353. MR Zbl

[Doreian and Stockman 1969] P. Doreian and N. Stockman, "A critique of the multidimensional approach to stratification", *Sociol. Rev.* **17**:1 (1969), 47–65.

[Engel 1988] U. Engel, "Status inconsistency and criss-cross in an adolescent society", *Int. Sociol.* **3**:3 (1988), 283–300.

[Evan and Simmons 1969] W. M. Evan and R. G. Simmons, "Organizational effects of inequitable rewards: two experiments in status inconsistency", *Adm. Sci. Quart.* **14**:2 (1969), 224–237.

[Galtung 1964] J. Galtung, "A structural theory of aggression", *J. Peace Res.* **1**:2 (1964), 95–119.

[Gorman 1968] W. M. Gorman, "The structure of utility functions", *Rev. Econ. Stud.* **35**:4 (1968), 367–390.

[Hartman 1974] M. Hartman, "On the definition of status inconsistency", *Amer. J. Sociol.* **80**:3 (1974), 706–721.

[Hernes 1969] G. Hernes, "On rank disequilibrium and military coups d'etat", *J. Peace Res.* **6**:1 (1969), 65–72.

[Hodge and Schwallier 2006] J. K. Hodge and P. Schwallier, "How does separability affect the desirability of referendum election outcomes?", *Theory Decis.* **61**:3 (2006), 251–276. MR Zbl

[Kriesberg 1998] L. Kriesberg, *Constructive conflicts: from escalation to resolution*, Rowman & Littlefield, Lanham, MD, 1998.

[Muller and Jukam 1983] E. N. Muller and T. O. Jukam, "Discontent and aggressive political participation", *Br. J. Political Sci.* **13**:2 (1983), 159–179.

[Pruitt and Kim 2004] D. G. Pruitt and S. H. Kim, *Social conflict: escalation, stalemate, and settlement*, McGraw Hill, New York, NY, 2004.

[Segal et al. 1970] D. R. Segal, M. W. Segal, and D. Knoke, "Status inconsistency and self-evaluation", *Sociom.* **33**:3 (1970), 347–357.

hodgejo@gvsu.edu   *Department of Mathematics, Grand Valley State University, Allendale, MI 49401, United States*

fayest@gmail.com   *Mount Holyoke College, South Hadley, MA 01075, United States*

jamie.woelk@gmail.com   *Western State Colorado University, Gunnison, CO 81231, United States*

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

2017   vol. 10   no. 1