involve

a journal of mathematics

msp

2015 vol. 8, no. 5

involve

msp.org/involve

MANAGING EDITOR

Kenneth S. Berenhaut, Wake Forest University, USA, berenhks@wfu.edu

PRODUCTION

Silvio Levy, Scientific Editor

Cover: Alex Scorpan

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2015 is US $140/year for the electronic version, and $190/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.721

A simplification of grid equivalence

Nancy Scherich

(Communicated by Kenneth S. Berenhaut)

In the work of Cromwell and Dynnikov, grid equivalence is given by the grid moves commutation, (de-)stabilization and cyclic permutation. This paper gives a proof that cyclic permutation is a sequence of (de-)stabilization and commutation grid moves.

1. Introduction

A *grid diagram* is a two-dimensional square grid such that each square within the grid is decorated with an $\times$, $\circ$ or is left blank. This is done in a manner such that every column and every row has exactly one $\times$ and one $\circ$ decoration. The *grid number* of a grid diagram is the number of columns (or rows) in the grid. See Figure 1 for an example. This paper follows the grid notation used by Manolescu, Ozsváth, Szabó and Thurston [Manolescu et al. 2007] (see also [Manolescu et al. 2009]) with the convention that the rows and columns are numbered top to bottom and left to right, respectively.

A grid diagram is associated with a knot, or link, by connecting the $\times$ and $\circ$ decorations in each column and row by a straight line with the convention that vertical lines cross over horizontal lines. These lines form strands of the knot, and removing the grid leaves a projection of the knot. As a result, grid diagrams represent particular planar projections of knots, or links. This process is illustrated in Figure 2. The *knot type of a grid* is the knot type of the knot associated with the grid.

It is important to note that the $\times$ and $\circ$ decorations can specify an orientation of the knot, but more importantly they mark the end points of the strands of the knot in that column or row. So, if two grid diagrams are the same up to opposite labeling of the $\times$ and $\circ$ decorations, then the grid diagrams are considered the same even though the labeling might suggest opposite orientations. Also, because grid diagrams are square, any result established for the columns of a grid is also understood for the rows by rotating the grid by 90 degrees, and vice versa.

MSC2010: 57M27.
Keywords: knot theory, grid diagrams.

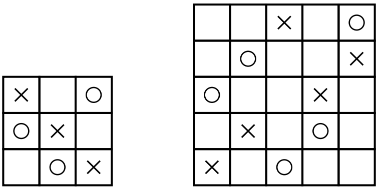


Figure 1. Grid diagrams with grid numbers 3 (left) and 5 (right).

There are three grid moves used to relate grid diagrams: commutation, cyclic permutation and (de-)stabilization. These play a role analogous to the Reidemeister moves [1932] for knot diagrams. Following the notation from [Manolescu et al. 2007], the three grid moves are as follows:

(1) Commutation interchanges two consecutive rows or columns of a grid diagram. This move preserves the grid number, as shown in Figure 3. Even though commutation may be defined for any two consecutive rows or columns, it is only permitted if the commutation preserves the knot type of the grid; refer to Section 2 for details. Throughout the introduction, it is assumed that all commutations preserve the knot type.

(2) Cyclic permutation preserves the grid number and removes an outer row/column and places it on the opposite side of the grid. See Figure 4.

(3) The third grid move has two different names depending on how the move is being used. Stabilization is the addition of a kink while destabilization is the removal. It is important to note that (de-)stabilization does not preserve the grid number. A kink may be added to the right or left of a column, and above or below a row. To

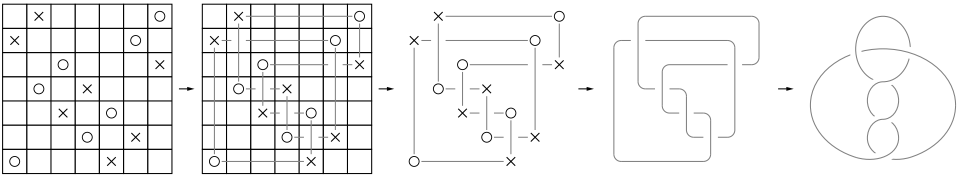


Figure 2. The process of finding the knot associated to a given grid diagram.

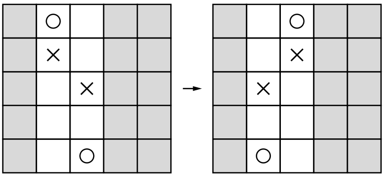


Figure 3. An example of column commutation.

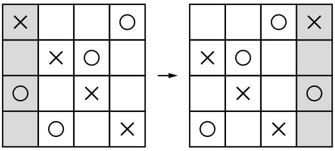


Figure 4. An example of column permutation.

add a kink to column c, insert an empty row between the $\times$ and $\circ$ markers of the column c. Then insert an empty column to the right or left of column c. Move either the $\times$ or $\circ$ decoration in column c into the adjacent grid square in the added column. Complete the added row and column with $\times$ and $\circ$ decorations appropriately. See Figure 5. To add a kink to a row, switch the notions of column and row. To remove a kink, follow these instructions in reverse order. As shown, stabilization increases the grid number by 1 while destabilization reduces the grid number by 1.

The following theorem explicates the relationship between grid diagrams, knots and the three grid moves.

Theorem 1.1 [Cromwell 1995; Dynnikov 2006]. *Let G_1, G_2 be a grid diagrams representing knots K_1, K_2 respectively. Then K_1 and K_2 are equivalent knots if and only if there exists a sequence of commutation*, (*de-*)*stabilization and cyclic permutation grid moves to relate G_1 to G_2.*

In other words, the three grid moves form an equivalence relation on the set of grid diagrams, and two grid diagrams are equivalent if and only if they represent the same knot. The three grid moves play a role similar to the Reidemeister moves [1932] for knot diagrams.

Grid diagrams have become increasingly widespread since the use of grids to give a combinatorial definition of knot Floer homology [Manolescu et al. 2007]. From the approach of knot Floer homology, invariance under cyclic permutation is trivial when viewed as diagrams on a torus. However, this paper will show that in any context, cyclic permutation is an unnecessary hypothesis of Theorem 1.1. In other words, the equivalence given by Theorem 1.1 can be strengthened so that two grid diagrams

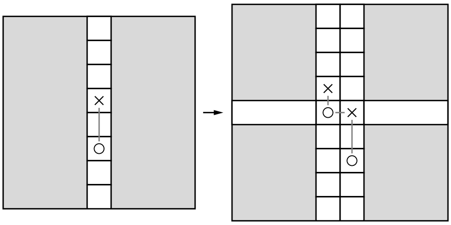


Figure 5. An example of stabilization, or kink addition.

are equivalent if there exists a sequence of commutation and (de-)stabilization grid moves to relate the two grid diagrams. This implies that invariance for any object defined using grids may be confirmed by checking invariance under only two moves: commutation and (de-)stabilization. This strengthened equivalence of grid diagrams is an immediate corollary to the following theorem.

Theorem 1.2. *There exists a sequence of commutation and* (*de-*)*stabilization grid moves that perform the cyclic permutation grid move.*

Corollary 1.3. *Let G_1, G_2 be grid diagrams representing knots K_1, K_2 respectively. Then K_1 and K_2 are equivalent knots if and only if there exists a sequence of commutation and* (*de-*)*stabilization grid moves to relate G_1 to G_2.*

The result of Theorem 1.2 is well known to certain experts. For example, the computer implementation of knot Floer homology available as part of KnotTheory`[1], due to Jean-Marie Droz, makes use of such a simplification. More concretely, after completing this project the author learned that Theorem 1.2 is proved in the work of Ozsváth, Szabó and Thurston [Ozsváth et al. 2008, Lemma 4.3]. However, since Theorem 1.2 is an interesting result in combinatorial knot theory in its own right, an independent proof is of value. Further, an illustrated proof of Theorem 1.2 may serve as a useful introduction to grid diagrams. The main goal of this paper is to provide a constructive proof of Theorem 1.2.

Organization of the paper. To prove Theorem 1.2, Section 2 addresses a subtlety of the commutation grid move required to preserve grid equivalence. Section 3 introduces four intermediate grid moves that when applied sequentially perform a cyclic permutation in terms of commutations and (de-)stabilizations. Lastly, Section 4 formalizes the proof of Theorem 1.2.

Terminology. The word *grid* will be used synonymously with *grid diagram* throughout the paper.

2. Commutation in detail

The commutation grid move is defined to interchange any two consecutive rows or columns in a grid. However, in some instances, commutation does not preserve the knot type of the grid. Since grids are useful as representations of knots with an equivalence relation generated by the grid moves, it is important to identify the exact conditions under which commutation preserves this equivalence relation. These conditions will be established for column commutation.

Figure 6 shows the four possible relative positions of two consecutive columns, up to different × and ○ labeling and exact spacing. Denote these possibilities as *nonshared, total-shared, partial-shared and point-shared*, see Figure 6.

[1] *KnotTheory`* is a Mathematica package and is available from www.katlas.org.

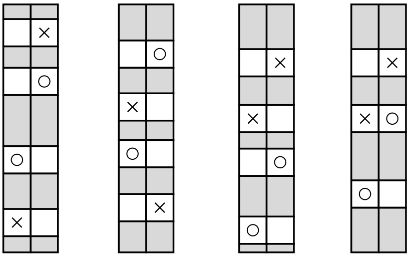


Figure 6. From left to right: nonshared, total-shared, partial-shared, and point-shared columns.

Lemma 2.1. *Commutation of nonshared, total-shared and point-shared columns preserves the knot type of the grid diagram.*

Proof. To prove these conditions preserve the knot type, consider the knot associated with the grid. The following will show that the associated knot is only altered by a Reidemeister I move, a Reidemeister II move or isotopy, thus preserving the knot equivalence class.

For nonshared columns, there are three scenarios, all resulting in isotopy. See Figure 7. For total-shared, there are three scenarios, two resulting in a Reidemeister II move and the other in isotopy. See Figure 8.

For point-shared there are four scenarios, two resulting in isotopy and two resulting in a Reidemeister I move. See Figure 9. □

Corollary 2.2. *Commutation of a column that has* × *and* ○ *decorations in adjacent grid squares will preserve the knot type of the grid.*

Proof. This column will only be nonshared, point-shared or total-shared with a consecutive column. □

Corollary 2.3. *Commutation of a column that has* × *and* ○ *decorations in the top and bottom grid squares will preserve the knot type of the grid.*

Proof. This column will always be total-shared with any consecutive column. □

Remark 2.4. Commutation of columns that are partial-shared may change the knot type of the grid.

Figure 10 shows two scenarios of partial-shared columns that, when commuted, change the crossings of the knot associated with the grid in a complicated way. The left scenario shows two strands that are not linked but become linked after the column commutation. The right shows how commutation changes an over-crossing to an under-crossing. In both of these scenarios, more knowledge about the knot would be needed to determine if the knot type was preserved.

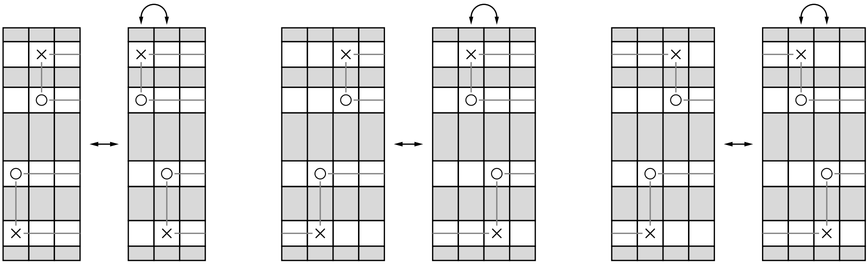


Figure 7. Three scenarios for nonshared columns.

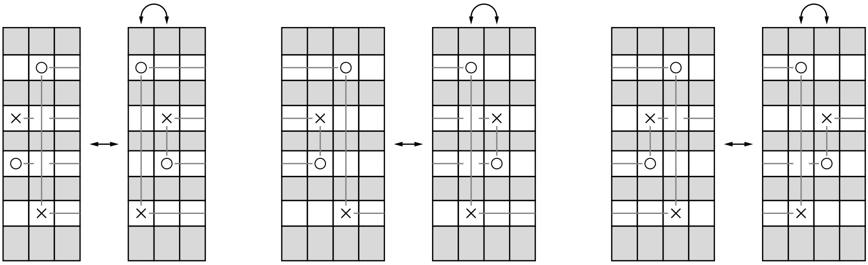


Figure 8. Three scenarios for total-shared columns. The left and middle result in a Reidemeister II move, and the right in isotopy.

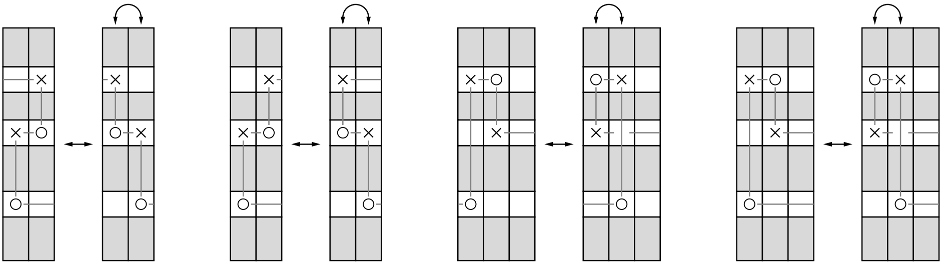


Figure 9. Four scenarios for point-shared columns. From left to right, the first two result in isotopy, and the second two in a Reidemeister I move.

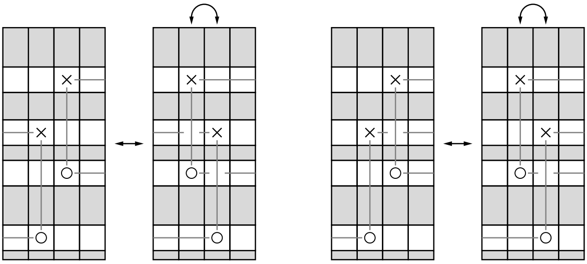


Figure 10. Partial-shared columns.

Remark 2.5. Point-shared commutation is not considered a standard grid move. In fact, it can be accomplished by a single destabilization followed by a single stabilization. This paper considers point-shared commutation with the sole interests of simplifying the proof of Corollary 2.2 and exhibiting a Reidemeister I move via grid moves in Figure 9. Often in the literature, namely [Manolescu et al. 2007] and [Ozsváth et al. 2008], point-shared commutation is not considered an allowable grid move. Throughout the remainder of the paper, point-shared commutation will not be used and the main result does not require this type of commutation.

3. Intermediate grid moves

The goal of the intermediate grid moves is to accomplish a column permutation from left to right using only commutations and (de-)stabilizations. A column permutation preserves the size of the grid and relative positions of the × and ○ decorations in the permuted column. So throughout the construction of the intermediate moves, any change in grid size or relative positioning of the × and ○ decorations in the permuted column will be noted.

The intermediate grid moves are independent from each other, but to simplify the proof of Theorem 1.2, each intermediate move will be described starting from the ending position of the previous intermediate move. Thus, when applied sequentially, it will be clear that a cyclic permutation is accomplished.

The first intermediate grid move I_1.

Definition 3.1. The I_1 move increases the grid number by 2 and moves the × and ○ decorations in the first column to occupy the top and bottom grid squares of the first column, as shown in Figure 11.

Proposition 3.2. *The I_1 move can be accomplished by a sequence of commutation and* (*de-*)*stabilization moves that preserve the grid equivalence class.*

Proof. Fix a grid diagram with grid number n. Assume that the row containing the × in the first column is above the row containing the ○. For alternate labeling, switch the roles of the × and ○. Let × be in row m and ○ be in row k with standard

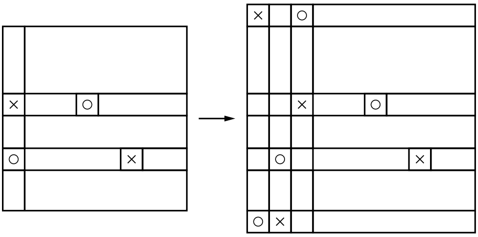
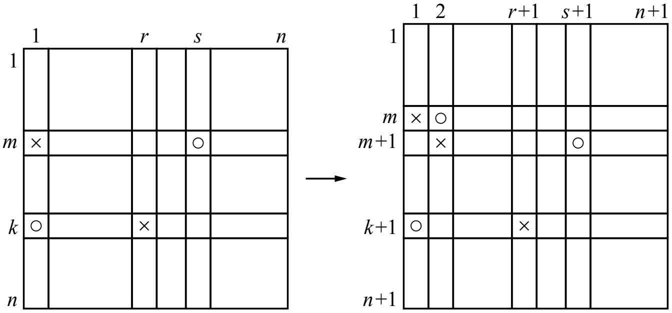


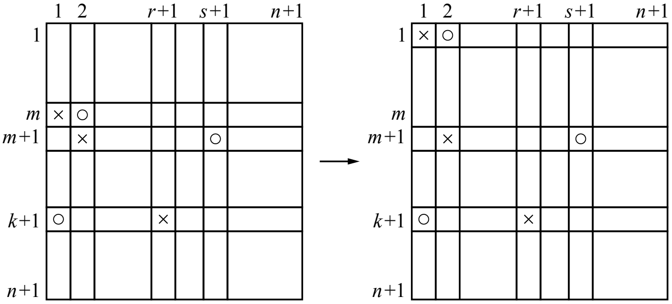
Figure 11. An illustration of the I_1 move.

top to bottom labeling. Let the ○ in the m-th row be in column s and the × in the k-th row be in column r.

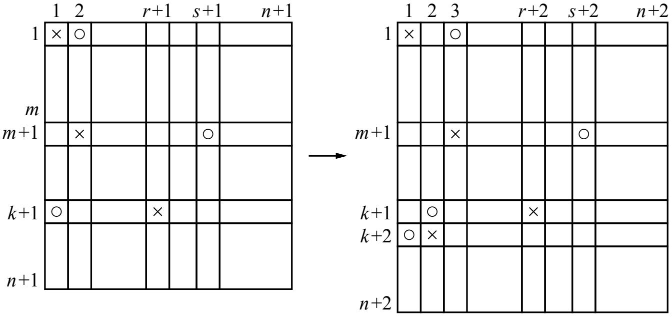
(1) Start by adding a kink above the m-th column. This increases the grid size by 1, resulting in a grid number of $n+1$.



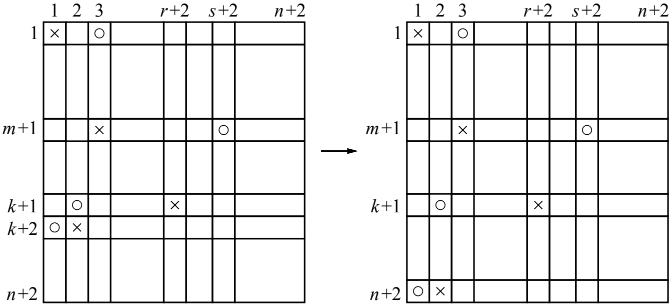
(2) The × and ○ in the first and second columns of row m are adjacent. By Corollary 2.2, commutation of row m preserves the grid equivalence class. So commute the row m upwards $m-1$ times making the × in the first column occupy the top row.



(3) After adding the kink, the ○ in the first column has been shifted down one row moving the ○ to the $(k+1)$-th row. Add a kink above the $(k+1)$-th row, moving the ○ to the $(k+2)$-th row. This increases the grid number by 1, resulting in a grid number of $n+2$.



(4) The ○ and × in the $(k{+}2)$-th row are adjacent, so by Corollary 2.2, commuting row $k{+}2$ preserves the knot type. Commute the $(k{+}2)$-th row downwards $(n+2)-(k+2)$ times until the ○ in the first column is in the bottom row.



Now the grid number increased to $n+2$ and the × and ○ decorations in the first column occupy the top and bottom grid squares of the first column. Since all commutations preserved the knot type, the grid equivalence class was preserved. □

The second intermediate move I_2.

Definition 3.3. Starting from the ending position of the I_1 move, where the × and ○ decorations in the first column occupy the top and bottom grid squares, the I_2 move cyclically permutes the first column to become the last column of the grid. (This is a special case of cyclic permutation). The I_2 move preserves the grid number. This is shown in Figure 12.

Proposition 3.4. *The I_2 move can be accomplished in a series of commutation grid moves and preserves the grid equivalence class.*

Proof. Since the × and ○ decorations in the first column are in the top and bottom grid squares, by Corollary 2.3, commutation of this column preserves the knot type of the grid. So commute the first column to the right $n{-}1$ times until it

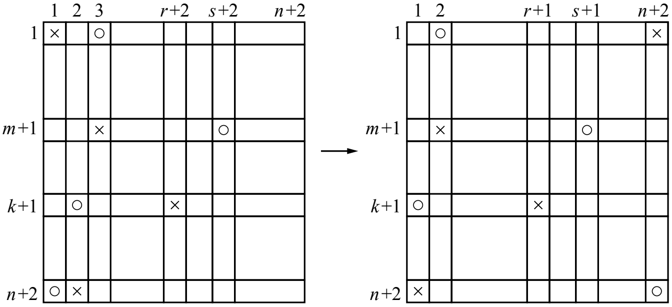


Figure 12. An illustration of the I_2 move.

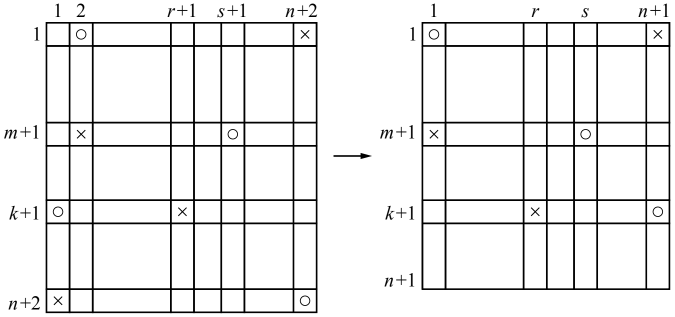


Figure 13. An illustration of the I_3 move.

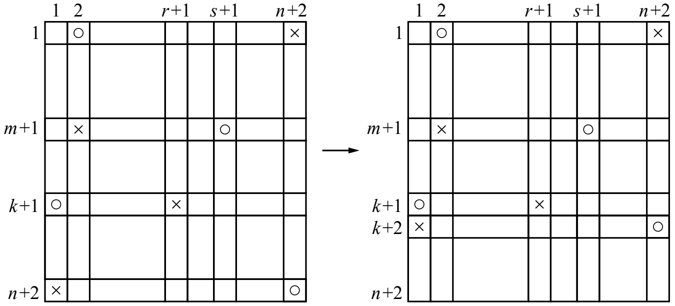
becomes the outermost right column. This clearly preserves the grid number and grid equivalence class. □

Third intermediate move I_3.

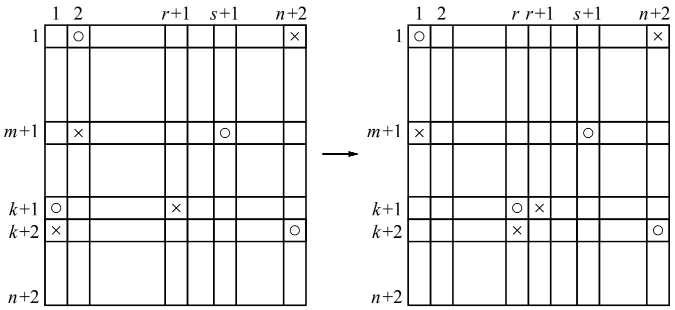
Definition 3.5. Starting from the ending position of the I_2 move, the move I_3 reduces the grid number by 1 and simplifies the bottom portion of the grid as shown in Figure 13.

Proposition 3.6. *The I_3 move can be accomplished by a sequence of commutation and (de-)stabilization grid moves and preserves the grid equivalence class.*

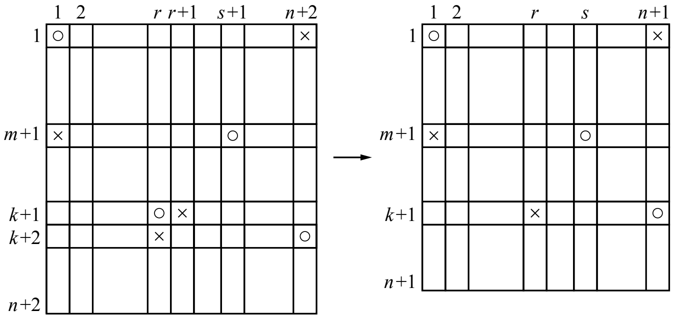
Proof. (1) Since the × and ○ decorations in the $(n+2)$-th row occupy the first and last grid squares, by Corollary 2.3 commuting this row preserves the knot type. So, commute the $(n+2)$-th row upwards $(n+2)-(k+1)-1$ times, until the × and ○ decorations in the $(k+1)$-th and $(k+2)$-th rows in the first column are adjacent.



(2) Since the × and ○ decorations in the first column are in adjacent grid squares, by Corollary 2.2 commuting this column preserves the knot type. So commute the first column to the right $r-1$ times until the × and ○ decorations in the $(k+1)$-th row are adjacent.



(3) Remove the kink in the r-th column and the $(n+2)$-th row, reducing the grid number to $n+1$.



Since all commutations preserved the knot type, the grid equivalence class was preserved and the grid number was reduced to $n+1$. □

Fourth intermediate move I_4.

Definition 3.7. Starting from the ending position of the I_3 move, the move I_4 mirrors the move I_3 and decreases the grid number to n as shown in Figure 14.

Proposition 3.8. *The I_4 move can be accomplished by a sequence of commutation and (de-)stabilization grid moves that preserve the grid equivalence class.*

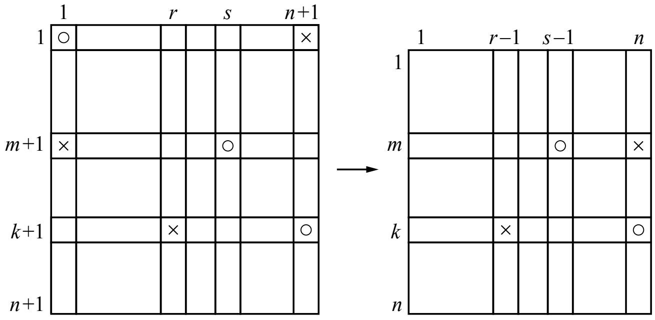
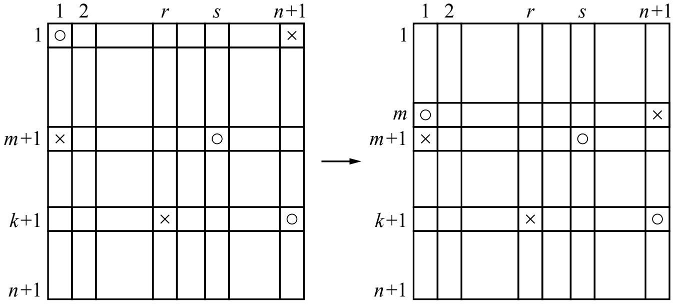
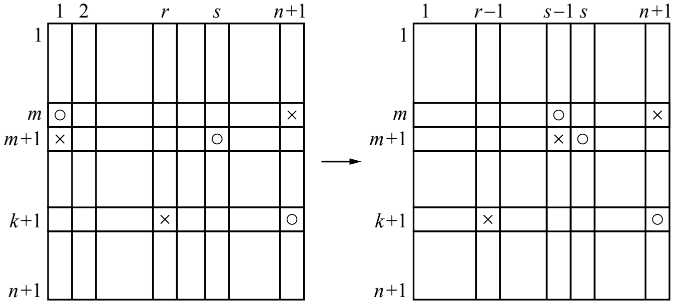


Figure 14. An illustration of the I_4 move.

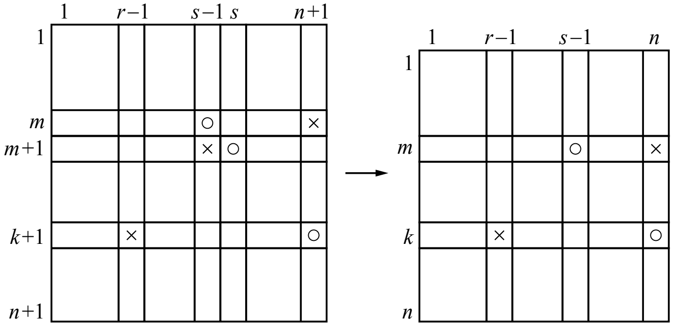
Proof. (1) Since the × and ○ decorations in the first row occupy the first and last grid squares, by Corollary 2.3 commuting this row preserves the knot type. So, commute the top row down $m-1$ times, so that the × and the ○ in the first column are adjacent.



(2) Since the × and ○ decorations in the first column are in adjacent grid squares, by Corollary 2.2 commuting this column preserves the knot type. So, commute the first column to the right s times until the × and ○ decorations in the $(m+1)$-th row are adjacent.



(3) Lastly, remove the kink in the $(s-1)$-th column and $(m+1)$-th row reducing the grid back to its original grid number n.



After the I_4 move, the grid number returns to the original value n, and the × and the ○ in the last column are in the same relative row positions as before the

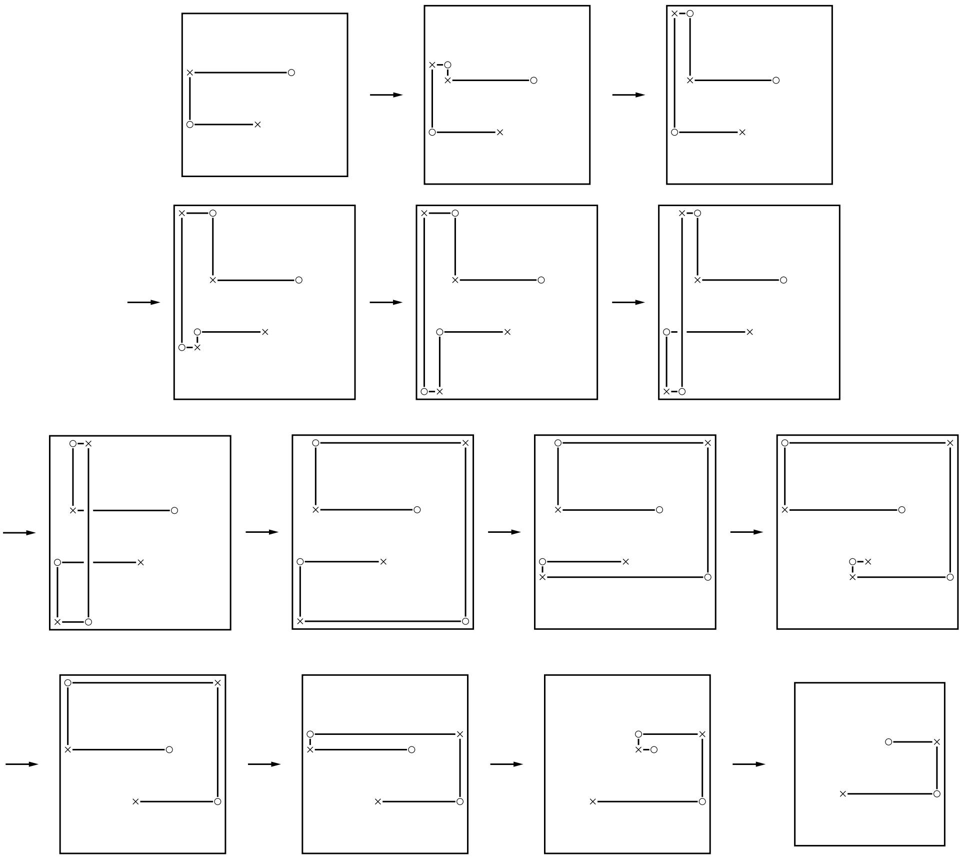


Figure 15. An illustration of the application of the intermediate grid moves used to produce a cyclic permutation grid move.

intermediate grid moves were applied. Since all commutations preserved the knot type, the grid equivalence class was preserved. □

4. Proof of Theorem 1.2

Theorem 1.2. *There exists a sequence of commutation and* (*de-*)*stabilization grid moves that perform the cyclic permutation grid move.*

Proof. Given a grid diagram, apply the intermediate grid moves I_1, I_2, I_3 and I_4 sequentially. As shown by construction, this sequence of intermediate moves preserves the grid number and relative row position of the $\times$ and $\circ$ decorations in the permuted column. Thus this sequence of intermediate moves performs a column permutation with only commutations and (de-)stabilizations. Figure 15 is a stylized diagram following the strand of the knot through the sequential application of the intermediate grid moves to explicate this construction. This process can be applied with an appropriate change of orientation to accomplish a cyclic permutation for a row or column in any direction. □

Acknowledgements

This paper formed part of a VIGRE funded undergraduate research project at UCLA (summer 2010). I would like to thank faculty advisor Liam Watson for significant guidance, inspiration, editing and choice of topic.

References

[Cromwell 1995] P. R. Cromwell, "Embedding knots and links in an open book, I: Basic properties", *Topology Appl.* **64**:1 (1995), 37–58. MR 96g:57006 Zbl 0845.57004

[Dynnikov 2006] I. A. Dynnikov, "Arc-presentations of links: monotonic simplification", *Fund. Math.* **190** (2006), 29–76. MR 2007e:57006 Zbl 1132.57006

[Manolescu et al. 2007] C. Manolescu, P. Ozsváth, Z. Szabó, and D. Thurston, "On combinatorial link Floer homology", *Geom. Topol.* **11** (2007), 2339–2412. MR 2009c:57053 Zbl 1155.57030

[Manolescu et al. 2009] C. Manolescu, P. Ozsváth, and S. Sarkar, "A combinatorial description of knot Floer homology", *Ann. of Math.* (2) **169**:2 (2009), 633–660. MR 2009k:57047 Zbl 1179.57022

[Ozsváth et al. 2008] P. Ozsváth, Z. Szabó, and D. Thurston, "Legendrian knots, transverse knots and combinatorial Floer homology", *Geom. Topol.* **12**:2 (2008), 941–980. MR 2009f:57051 Zbl 1144.57012

[Reidemeister 1932] K. Reidemeister, *Knotentheorie*, Ergebnisse der Mathematik und ihrer Grenzgebiete **1**, Springer, Berlin, 1932. Reprinted in 1974. MR 49 #9828 Zbl 0005.12001

Received: 2012-11-20 Revised: 2014-10-21 Accepted: 2015-01-16

nancy.scherich@gmail.com
Department of Mathematics, University of California, Santa Barbara, Santa Barbara, CA 93106, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.735

A permutation test for three-dimensional rotation data

Daniel Bero and Melissa Bingham

(Communicated by Mary C. Meyer)

Statistical inference procedures that require no distributional assumptions make up the area of nonparametric statistics. The permutation test is a common nonparametric test that can be used to compare measures of center for two data sets, but it is yet to be explored for three-dimensional rotation data. A permutation test for such data is developed and the statistical power of this test is considered under various scenarios. The test is then used in an application comparing movement around joints in the foot and ankle for humans, chimpanzees, and baboons.

1. Introduction

Data in the form of three-dimensional rotations are common in the study of human motion. As skeletal mammals move, the orientation of various joints can be tracked by using infrared emitting diodes attached to bones on opposite ends of the joint. Each joint orientation can be represented mathematically as a 3×3 orthogonal rotation matrix. Of interest here is comparing movement around various joints in the ankle and foot for humans, chimpanzees, and baboons by comparing the central rotation of each joint for the various species.

While other works have considered comparing sets of three-dimensional rotation data, they rely on distributional assumptions [Rancourt et al. 2000; Hendriks and Landsman 1998]. Further, existing work for studying three-dimensional rotations is often in terms of manifold considerations. As such, it is often inaccessible to practitioners outside the area. Our aim here is development of methodology for comparing central rotations that is both nonparametric and does not rely on special manifold theory, so that it can be used more broadly. The permutation test is a commonly used nonparametric test, but it has yet to be implemented for three-dimensional rotation data. We develop such a test in Section 2, explore the statistical power of the test in Section 3, and apply the test to joint data in Section 4.

MSC2010: 62G09, 62G10.

Keywords: nonparametric statistics, permutation test, directional data.

This research was supported by NSF grant DMS-1104409.

2. Development of a three-dimensional permutation test

The permutation test is widely used in nonparametric statistics for determining if two data sets are different in some way (e.g., comparing means, variances, shapes). The most common example of a permutation test in one dimension is comparing population means for data sets A and B by using the difference in sample means, $\bar{x}_A - \bar{x}_B$, as a test statistic. To perform the permutation test, data sets A and B are combined and permuted so that data points are randomly reassigned to either A or B. The permuted test statistic is then calculated from this permuted data and this process is repeated a large number of times. If the means of the populations from which A and B come do in fact differ, then we expect the observed test statistic $\bar{x}_A - \bar{x}_B$ to be more extreme than the permuted test statistics. For this reason, the p-value for a permutation test is defined to be the proportion of times that the permuted test statistic is more extreme than the observed test statistic. See [Higgins 2004] for more details on permutation tests.

To translate the idea of the permutation test to three-dimensional rotation data, we first need to define a sensible test statistic that could be used for comparing two central rotations. For each set of three-dimensional rotations, we begin by finding a measure of center as follows. Compute $\overline{\boldsymbol{O}} = 1/n \sum_{i=1}^{n} \boldsymbol{O}_i$ for $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_n \in \mathrm{SO}(3)$, where SO(3) represents the set of all 3×3 orthogonal rotation matrices. Next, find the matrix $\boldsymbol{T} = \boldsymbol{V}\boldsymbol{W}$, where $\overline{\boldsymbol{O}} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{W}$ is the singular value decomposition of $\overline{\boldsymbol{O}}$. Using these components from the singular value decomposition is necessary since $\overline{\boldsymbol{O}}$ may not be an element of SO(3), but $\boldsymbol{T}$ is. This is a commonly used measure of center [León et al. 2006; Bingham et al. 2009; Khatri and Mardia 1977], which we refer to as the "mean" rotation.

Once we have found the mean rotation for each of our two data sets, a natural test statistic is the difference between these mean rotations. One way of quantifying the difference between two three-dimensional rotations is by using angles. A misorientation angle is defined as the angle needed to rotate from one three-dimensional rotation to another via a spin about some axis. For $\boldsymbol{O}, \boldsymbol{P} \in \mathrm{SO}(3)$, the misorientation angle between $\boldsymbol{O}$ and $\boldsymbol{P}$ is

$$\mathrm{mis}(\boldsymbol{O}, \boldsymbol{P}) = \arccos\left(\frac{\mathrm{tr}(\boldsymbol{O}'\boldsymbol{P}) - 1}{2}\right), \tag{1}$$

where tr is the trace of a matrix and $\boldsymbol{O}'$ is the transpose of $\boldsymbol{O}$. We use the misorientation angle between our two mean rotations as the test statistic for the three-dimensional permutation test of H_o: There is no difference between the population mean rotations versus H_a. There is a difference between the population mean rotations. The steps of the permutation test are given below and R code for implementing this test is provided in the Appendix.

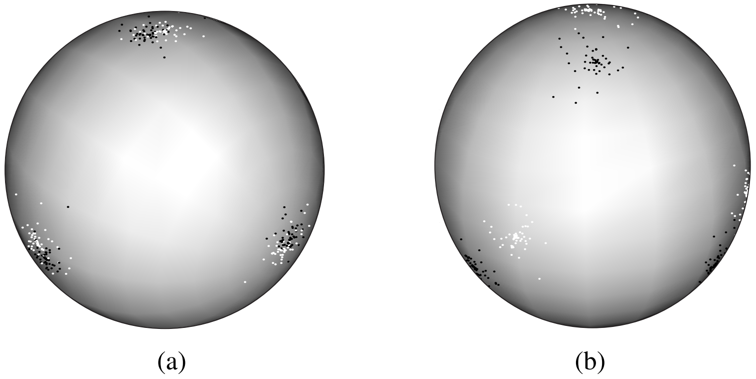


Figure 1. Plots of two simulated three-dimensional rotation data sets (each with $n = 50$) with mean rotations that (a) are not significantly different and (b) are significantly different.

(1) Calculate the mean rotation for each data set and then find the misorientation angle between these means. This serves as the observed test statistic, θ_{obs}.

(2) Permute the data a large number (say 10,000) of times, storing the misorientation angle between the permuted mean rotations, θ_{perm}, each time.

(3) Let the p-value be the fraction of times that the permuted misorientation angle is greater than the observed misorientation angle; that is,

$$p\text{-value} = \frac{\#\text{ of times } \theta_{\text{perm}} > \theta_{\text{obs}}}{\#\text{ of permutations}}.$$

The three-dimensional permutation test outlined above is briefly illustrated in two different examples. Figure 1 shows three-dimensional data sets plotted as points on the sphere, with one observation represented by three points that would correspond to three orthogonal axes. In Figure 1(a), the two simulated data sets (in white and black, each of size 50) show considerable overlap. Under the permutation test, these data sets resulted in a test statistic of 0.0546 and a p-value of 0.3101. In Figure 1(b), the simulated data sets are more separated. These data sets gave a test statistic of 0.6102 and a p-value of 0, indicating a significant difference in the population mean rotations. These examples suggest that the p-value decreases as expected when the data sets have mean rotations that increase in distance.

3. Power: a simulation study

To examine the effectiveness of the three-dimensional permutation test developed in Section 2, we perform a simulation study to investigate statistical power. Power is the probability of correctly rejecting a false null hypothesis. We simulate data sets with centers that differ by a known misorientation angle, ϕ, (i.e., there is a

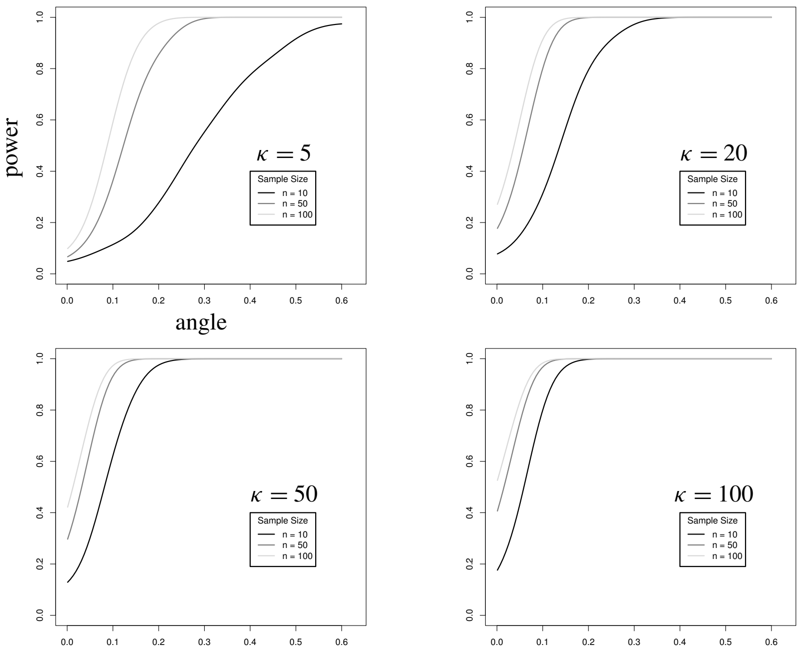


Figure 2. Plots of power versus misorientation angle for the von Mises version of the UARS distributions with $\kappa = 5, 20, 50, 100$.

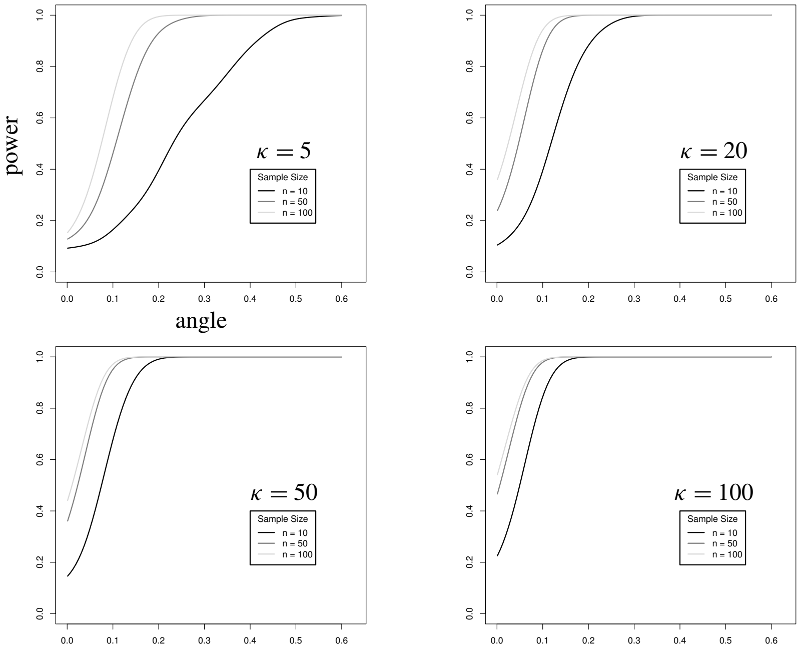


Figure 3. Plots of power versus misorientation angle for the symmetric matrix von Mises–Fisher distribution with $\kappa = 5, 20, 50, 100$.

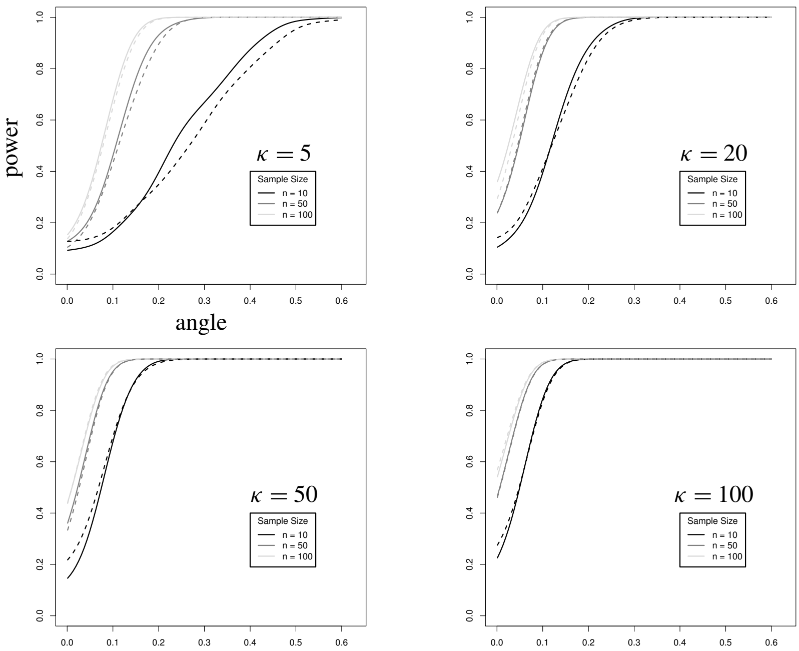


Figure 4. Plots of power versus misorientation angle for the vM-F distribution with solid lines representing the permutation test and dashed lines representing the parametric approach for $\kappa = 5, 20, 50, 100$.

difference between the population mean rotations and the null is false) from both the von Mises version of the uniform axis-random spin (vM-UARS) distribution [Bingham et al. 2009] and the symmetric version of the matrix von Mises–Fisher (vM-F) distribution [Khatri and Mardia 1977] . A vM-UARS or vM-F distribution can be specified by a central rotation $\boldsymbol{S} \in \mathrm{SO}(3)$ and a spread parameter $\kappa \in (0, \infty)$, where κ is best termed as a concentration parameter since larger values of κ indicate rotations that are less spread about the center at $\boldsymbol{S}$. Two samples, each of size n, are generated from vM-UARS$(\boldsymbol{S}_1, \kappa)$ and vM-UARS$(\boldsymbol{S}_2, \kappa)$ distributions, where $\phi = \mathrm{mis}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ as in (1). We consider κ values of 5, 20, 50, and 100, set n at 10, 50, and 100, and let the misorientation angle, ϕ, vary between 0 and $\pi/5$. The same is done for the vM-F distribution.

For each combination of κ, n, and ϕ, the permutation test was conducted 1,000 times with 1,000 permutations per test. The power was then found as the proportion of times (out of 1,000) that the test correctly rejected the null hypothesis of equal means. Plots of the power against the misorientation angle, ϕ, for the various choices of n and κ are provided in Figure 2 for the vM-UARS distribution and in Figure 3 for the vM-F distribution. It can be seen from all plots that as sample size increases,

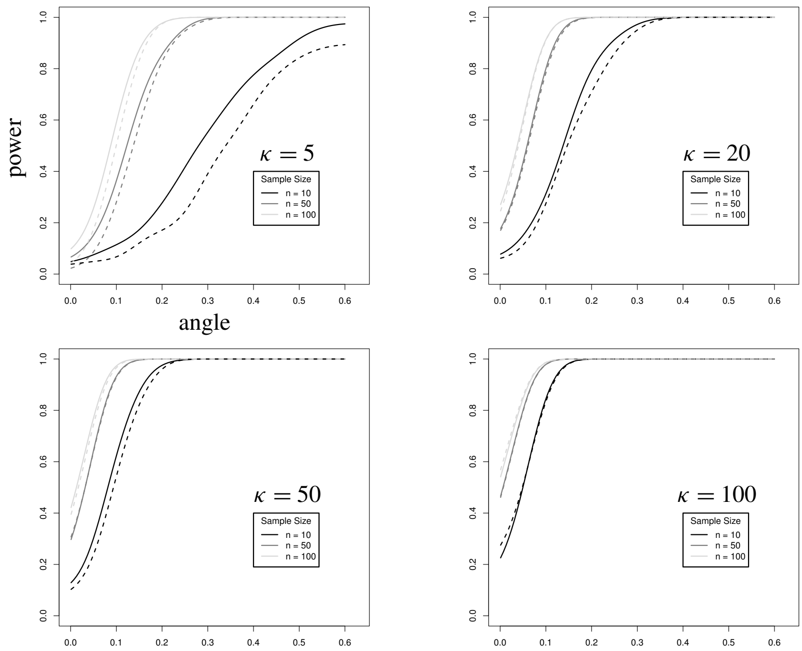


Figure 5. Plots of power versus misorientation angle for the vM-UARS distribution with solid lines representing the permutation test and dashed lines representing the parametric approach for $\kappa = 5, 20, 50, 100$.

the power of the test increases. In addition, as the concentration parameter, κ, increases (i.e., data sets become more clustered around their mean rotation), the power increases. Finally, as the misorientation angle increases and the true centers become farther apart, the power increases. This mimics properties of power for traditional hypothesis tests for differences in means (for nonrotational data), giving evidence that the three-dimensional permutation test performs as desired.

The power of the three-dimensional permutation test was also compared to that of the parametric approach presented in [Rancourt et al. 2000], which requires the observations be distributed according to the matrix von Mises–Fisher distribution. The plots in Figure 4 show power versus misorientation angle for the various choices of n and κ using the matrix von Mises–Fisher distribution. The solid lines represent power for the permutation test, with the dashed lines representing power for the parametric approach. We see that the power of the permutation test is comparable to the power of parametric approach in all cases. The permutation test was also compared to the parametric approach for the vM-UARS distribution, with power plots given in Figure 5. We see that the permutation test outperforms the parametric approach in terms of resulting in a larger power, with this fact more visible when

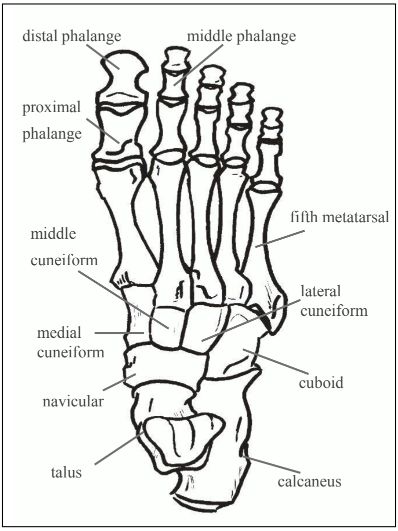


Figure 6. Bones in the ankle and foot (image taken from http://www.ceuarmy.com/BSFAFpdf.pdf).

we have smaller sample sizes or data that is more spread (small κ). Thus, the three-dimensional permutation test is comparable to the parametric approach when the assumptions of the parametric test are met, and it performs better than the parametric approach when the assumptions are not met.

4. Application to ankle joint rotation data

Now that we have verified that the three-dimensional permutation test performs as expected with regard to power, we apply the test to ankle/foot joint rotation data collected by Prof. Thomas Greiner of the Department of Physical Therapy at the University of Wisconsin-La Crosse. Data was collected from humans, baboons, and chimps during circumduction, which is the movement characterized by the foot being placed flat on the floor and the leg rotating in a circular motion around it. Infrared emitting diodes attached to bones on each side of a joint give the orientation of each bone as the movement occurs. If the orientation of the first bone is represented as $\boldsymbol{F}$ and the orientation of second bone is represented as $\boldsymbol{G}$, then the resulting orientation of the joint is defined as $\boldsymbol{F}'\boldsymbol{G}$. Because markers may not have been placed identically on all subjects, the orientations of all joints under consideration were measured with the tibia-talus joint as the reference to allow for comparison of species. Joints considered were the cuboid-calcaneus, navicular-cuboid, navicular-talus, talus - calcaneus, and fifth metatarsal-cuboid. (See Figure 6 for a diagram of the bones in the foot and ankle region.) Orientations were collected for six human subjects, four chimpanzee subjects, and seven baboon subjects, and the base alignment matrix corresponding to the primary rotational axis (see [Ball and Greiner 2012]) was used in the three-dimensional permutation test to compare species.

Species were compared pairwise (human versus chimpanzee, human versus baboon, and chimpanzee versus baboon) for each of the joints mentioned above, and each test was done using 1,000 permutations. Out of all tests, there were four significant differences found. There was significant evidence to suggest that the orientation of the navicular-talus joint differs between the humans and chimpanzees (p-value = 0.001) and humans and baboons (p-value $\approx$ 0). The orientation of the talus-calcaneus joint was found to be significantly different between humans and chimpanzees (p-value = 0.019) and humans and baboons (p-value = 0.001). Therefore, it appears that movement for humans differs from baboons and chimps when considering two specific joints.

5. Conclusion

The analysis of joint rotation data provided here is just one of many applications that the three-dimensional permutation test could be used for. Given the abundance of three-dimensional rotation data in the study of human motion, as well as in the other fields like materials science, having methodology for comparing measures of center for three-dimensional data is important. The three-dimensional permutation test developed here provides that methodology without the need for any distributional assumptions on where the data sets come from. It also does not require any theory on special manifolds, making the three-dimensional permutation test an important addition to the field of statistics, as well as to practitioners who collect data in this form.

Appendix

The following gives an R function called `PermTest` for performing the three-dimensional permutation test on data sets A (of size n_A) and B (of size n_B). The argument `A` must be an array of dimension $3 \times 3 \times n_A$ and `B` must be an array of dimension $3 \times 3 \times n_B$. The argument `nspec` specifies the number of times the data should be permuted. The function `PermTest` outputs the test statistic (misorientation angle between the two sample mean rotations) and p-value.

```
PermTest=function(A,B,nspec){
  ##Loads functions needed for test
  trace=function(M){sum(diag(M))}
  Mis.Ang=function(C,D){acos((trace(t(C)%*%D)-1)/2)}

  ##Finds mean matrices for both sets of data
  na=dim(A)[3]
  Abar=matrix(rep(0,9),nrow=3)
  for(i in 1:na){Abar=Abar+A[,,i]}
  Abar=Abar/na
  M.A=svd(Abar)$u%*%t(svd(Abar)$v)
  nb=dim(B)[3]
  Bbar=matrix(rep(0,9),nrow=3)
  for(i in 1:nb){Bbar=Bbar+B[,,i]}
  Bbar=Bbar/nb
  M.B=svd(Bbar)$u%*%t(svd(Bbar)$v)

  ##Finds the test statistic
  Test.Stat=Mis.Ang(M.A,M.B)

  ##Puts data into one array
  T=array(c(A,B),dim=c(3,3,(na+nb)))

  ##Performs the permutation test
  nsim=nspec
  ang=rep(0,nsim)
  for(i in 1:nsim){
    samp=sample(1:(na+nb))
    O=T[,,samp[1:na]]
    P=T[,,samp[(na+1):(na+nb)]]
    Obar=matrix(rep(0,9),nrow=3)
    for(j in 1:na){Obar=Obar+O[,,j]}
    Obar=Obar/na
    M.O=svd(Obar)$u%*%t(svd(Obar)$v)
    Pbar=matrix(rep(0,9),nrow=3)
    for(k in 1:nb){Pbar=Pbar+P[,,k]}
    Pbar=Pbar/nb
    M.P=svd(Pbar)$u%*%t(svd(Pbar)$v)
    ang[i]=Mis.Ang(M.O,M.P)
    }
  p.value=sum(ang>Test.Stat)/nsim
  list(Test.Statistic=Test.Stat,P.Value=p.value)
}
```

References

[Ball and Greiner 2012] K. A. Ball and T. M. Greiner, "A procedure to refine joint kinematic assessments: functional alignment", *Comput. Methods Biomech. Biomed. Eng.* **15**:5 (2012), 487–500.

[Bingham et al. 2009] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Modeling and inference for measured crystal orientations and a tractable class of symmetric distributions for rotations in three dimensions", *J. Amer. Statist. Assoc.* **104**:488 (2009), 1385–1397. MR 2011a:62189 Zbl 1205.62215

[Hendriks and Landsman 1998] H. Hendriks and Z. Landsman, "Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions", *J. Multivariate Anal.* **67**:2 (1998), 227–243. MR 2000a:62125 Zbl 0941.62069

[Higgins 2004] J. J. Higgins, *Introduction to modern nonparametric statistics*, Brooks/Cole, Pacific Grove, CA, 2004.

[Khatri and Mardia 1977] C. G. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics", *J. Roy. Statist. Soc.* (*B*) *Stat. Methodol.* **39**:1 (1977), 95–106. MR 58 #13506 Zbl 0356.62044

[León et al. 2006] C. A. León, J.-C. Massé, and L.-P. Rivest, "A statistical model for random rotations", *J. Multivariate Anal.* **97**:2 (2006), 412–430. MR 2234030 Zbl 1085.62066

[Rancourt et al. 2000] D. Rancourt, L.-P. Rivest, and J. Asselin, "Using orientation statistics to investigate variations in human kinematics", *J. Roy. Statist. Soc.* (*C*) *Appl. Stat.* **49**:1 (2000), 81–94. MR 1817876 Zbl 0974.62107

Received: 2013-06-03 Revised: 2014-07-22 Accepted: 2014-09-07

dbero@iastate.edu *Colony Brands, Monroe, WI 53566, United States*

mbingham@uwlax.edu *Mathematics Department, University of Wisconsin-La Crosse,*
1725 State Street, La Crosse, WI 54601, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.745

Power values of the product of the Euler function and the sum of divisors function

Luis Elesban Santos Cruz and Florian Luca

(Communicated by Filip Saidak)

We find examples of positive integers n such that $\phi(n^3)\sigma(n^3)$ is a perfect square.

1. Introduction

The Euler function $\phi(n)$ counts the number of positive integers $m \le n$ which are coprime to n, the sum of divisors function $\sigma(n)$ is equal to the sum of the positive proper divisors of n, and both of these functions have fascinated mathematicians for centuries. A lot of effort has been spent trying to find positive integers n such that $\phi(n)$ and $\sigma(n)$ have nice arithmetic properties.

It is easy to make $\phi(n)$ a square. Just take $n = 2^{2k+1}$ for some $k \ge 0$. Exactly half of all integers $m \le 2^{2k+1}$ are odd, and hence, coprime to n. Thus, $\phi(2^{2k+1}) = 2^{2k}$ is a perfect square. The situation for the sum of divisors function is harder. A nice presentation of this problem is in [Beukers et al. 2012]. Following that reference, we look at the factorizations

$$\begin{aligned} \sigma(2) &= 3, & \sigma(11) &= 2^2 \times 3, \\ \sigma(3) &= 2^2, & \sigma(13) &= 2 \times 7, \\ \sigma(5) &= 2 \times 3, & \sigma(17) &= 2 \times 3^2, \\ \sigma(7) &= 2^3, & \sigma(19) &= 2^2 \times 5. \end{aligned}$$

There are many ways to multiply together some of the above numbers to get a perfect square. First let us notice that 13 and 19 are useless because $\sigma(13) = 2 \times 7$ and $\sigma(19) = 2^2 \times 5$, and neither 7 nor 5 ever appear again on the right-hand side of the above equations. Throw out 13 and 19 and group squares on the right-hand sides in the following way, where $\square$ represents a perfect square:

$$\sigma(2) = 3, \quad \sigma(3) = \square, \quad \sigma(5) = 2 \times 3, \quad \sigma(7) = 2\square, \quad \sigma(11) = 3\square, \quad \sigma(17) = 2\square.$$

MSC2010: 11B68, 11A25.
Keywords: sum of divisors, Euler function.
Santos Cruz worked on this paper during a summer project under of the supervision of Luca.

Note that all six inputs are prime numbers and all outputs have prime factorizations consisting of only 2 and 3. Let the primes 2, 3, 5, 7, 11, 17 correspond to the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5, \boldsymbol{v}_6$ in the six-dimensional vector space $\mathbb{F}_2^6$, where $\boldsymbol{v}_i$ has i-th component equal to 1 and all others equal to 0 for $i = 1, \ldots, 6$. In $\mathbb{F}_2^2$ we let $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ be the vectors $(1, 0)^\top$ and $(0, 1)^\top$ and think of them as corresponding to the primes 2 and 3 respectively. We define a linear map from $\mathbb{F}_2^6 \mapsto \mathbb{F}_2^2$ whose matrix is

$$T = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

This matrix has rank 2, so it has $2^4 = 16$ vectors in its nullspace, and any of these vectors gives us a solution. For example, the vector $(1, 1, 1, 1, 0, 0)^\top$, which is in $\mathrm{Null}(T)$, gives us the solution $n = 2 \times 3 \times 5 \times 7$, having $\sigma(n) = 2^6 \times 3^2$.

In [Beukers et al. 2012], the equation $\sigma(n^k) = m^l$ in positive integers n and m was studied for some exponents $k > 1$ and $l > 1$. On page 377, they conjecture that $\sigma(n^k) = m^l$ has only finitely many solutions if $k > 3$ and $l > 1$ are given. Here, we propose the following counterconjecture.

Conjecture 1. *For every $k > 1$ and $l > 1$, there are infinitely many n such that $\sigma(n^k) = m^l$ for some positive integer m.*

To give some evidence, we propose a different conjecture. Let $P(n)$ denote the largest prime factor of the integer n, with the convention that $P(0) = P(\pm 1) = 1$.

Conjecture 2. *Let $f(x) \in \mathbb{Z}[x]$ be a polynomial such that $f(0) \neq 0$. For every $\varepsilon > 0$, there exists $c := c(\varepsilon)$ and $x_0 := x_0(\varepsilon)$ such that*

$$\#\{p \le x : P(f(p)) < x^\varepsilon\} \; > \; cx/\log x \quad \textit{for all} \quad x > x_0. \tag{1}$$

The substance of the above conjecture is the following. It is well known that the numbers n such that $P(n) < n^\varepsilon$ form a positive-density subset of $\mathbb{N}$. It is conjectured that the primes p such that $P(p-1) < p^\varepsilon$ form a positive-density subset of all primes. This is not known for small values of $\varepsilon > 0$. So, we venture even further and replace $p - 1$ by any fixed polynomial $f(p)$ such that $f(0) \neq 0$ (in order to make sure that p does not show up as a natural divisor of $f(p)$) and conjecture that, in fact, the set of primes p such that $P(f(p)) < p^\varepsilon$ is of positive density. This is known if all roots of $f(x)$ are rational, with some $\varepsilon < 1$ (like $\varepsilon = 1 - 1/2d$, where d is the degree of $f(x)$), but it is not known for any $\varepsilon < 1$ once $f(x)$ has an irreducible factor of degree at least 2. The quantity $x/\log x$ in the right-hand side of (1) arises from the prime number theorem, which asserts that, asymptotically, the function $\pi(x) = \#\{p \le x\}$ equals $x/\log x$ as $x \to \infty$.

Let us see how Conjecture 1 would follow from Conjecture 2. Let $k \ge 2$, $f(x) = (x^{k+1} - 1)/(x - 1)$ and suppose first that $l = 2$. Let x be large, put $\varepsilon = 1/2$

and let $p_1, \dots, p_t$ be such that $P(f(p_i)) < x^{1/2}$. Let $s = \pi(x^{1/2})$. Then we can write

$$f(p_i) = w_i \square, \quad i = 1, \dots, t,$$

where the w_i are square-free numbers with $P(w_i) \le x^{1/2}$. As before, we can identify the w_i with vectors in $\mathbb{F}_2^s$ obtained by putting 1 or 0 in the j-th component according to whether the j-th prime divides w_i or not. In this way, we get a linear application from $\mathbb{F}_2^t$ to $\mathbb{F}_2^s$ whose nullspace has dimension at least $t - s$, where

$$t - s > c\frac{x}{\log x} - \pi(x^{1/2}) > c\frac{x}{\log x} - x^{1/2},$$

and this last function certainly tends to infinity with x. This is when $l = 2$. Assume now that $l > 2$. Then we write

$$f(p_i) = w_i u_i^l \quad \text{for all} \quad i = 1, \dots, t,$$

where the w_i are l-th power free and $P(w_i) \le x^{1/2}$. We attach to each w_i an element $\boldsymbol{w}_i$ in the group $(\mathbb{Z}/l\mathbb{Z})^s$ where in the j-th component we put the exponent of the j-th prime number in the factorization of w_i. Note that $\mathbb{Z}/l\mathbb{Z}$ is not a field unless l is a prime, and even if l is a prime, we only can multiply *distinct* primes p_i in attempts to create n such that $\sigma(n^k) = m^l$. Thus, we are only allowed to take sums of distinct $\boldsymbol{w}_i$ and get 0. There is a theorem (see [van Emde Boas and Kruyswijk 1967] and [Olson 1969, Theorem 1]) that says that if we have at least $s(l-1)$ such distinct elements $\boldsymbol{w}_i$, we can find some of them whose sum is 0. Thus, we can create at least $\lfloor t/(s(l-1)) \rfloor$ distinct (in fact, even disjoint) subsets of the $\boldsymbol{w}_i$ for $i = 1, \dots, t$ simply by finding some 0-sum among the first $s(l-1)$ of them, another 0-sum among the next $s(l-1)$ of them and so on. Since

$$\frac{t}{s(l-1)} > \frac{c}{(l-1)} \frac{\sqrt{x}}{\log x},$$

and the right-hand side is a function that tends to infinity with x, we get Conjecture 1.

We can ask similar questions simultaneously for $\phi(n)$ and $\sigma(n)$, like making them simultaneously squares, or cubes, etc. This has already been treated in [Freiberg 2012]. There it is shown that the number of $n \le x$ such that both $\phi(n)$ and $\sigma(n)$ are perfect powers of an exponent l is less than $c_1 l x^{1/l}/(\log x)^{l+2}$, where $c_1 > 0$ is some positive constant. Square values of the product $\phi(n)\sigma(n)$ have been investigated in [Broughan et al. 2013]. In the next section, we present some computational examples of n such that $\phi(n^3)\sigma(n^3) = \square$.

2. Computational examples

We wanted to find a positive integer n such that $\phi(n^3)\sigma(n^3) = \square$. For a prime p, we have $\phi(p^3)\sigma(p^3) = p^2(p^4 - 1)$. So, we wrote $p^4 - 1 = w_p \square$, where w_p is square-free for all $p \le 1000$. Then we searched for a subset $\mathcal{S}$ of cardinality t such that the set of prime factors appearing in the factorizations of w_p for $p \in S$ has cardinality $s < t$. We found the subset

$$\{2, 3, 5, 7, 13, 17, 23, 31, 41, 43, 47, 73, 83, 191, 239, 307, 443, 499, 829\},$$

with $t = 21$ and $s = 17$. Thus, this set gives us $2^{21-17} = 16$ solutions. We wrote down the $\{0, 1\}$ matrix with 17 rows and 21 columns, which ends up having rank 17 over $\mathbb{F}_2$. The largest solution in the nullspace of this matrix is

$$n = 3 \times 7 \times 11 \times 13 \times \times 17 \times 23 \times 43 \times 47 \times 83 \times 239 \times 443 \times 499 \times 829,$$

for which $\phi(n^3)\sigma(n^3) = m^2$, where

$$m = 2^{30} \times 3^7 \times 5^{10} \times 7^2 \times 11 \times 13^4 \times 17^3 \times 23 \times 29 \times 37 \times 41 \times 53 \times 61 \times 83 \times 157.$$

Despite our efforts, we could not find an integer $n > 1$ such that $\sigma(n^5) = \square$, and we leave finding such an example as a challenge to the reader.

References

[Beukers et al. 2012] F. Beukers, F. Luca, and F. Oort, "Power values of divisor sums", *Amer. Math. Monthly* **119**:5 (2012), 373–380. MR 2916476 Zbl 1271.11088

[Broughan et al. 2013] K. Broughan, K. Ford, and F. Luca, "On square values of the product of the Euler totient and sum of divisors functions", *Colloq. Math.* **130**:1 (2013), 127–137. MR 3034320 Zbl 1286.11005

[van Emde Boas and Kruyswijk 1967] P. van Emde Boas and D. Kruyswijk, "A combinatorial problem on finite Abelian groups", *Math. Centrum Amsterdam Afd. Zuivere Wisk.* **1967**:ZW-009 (1967), 27. MR 39 #2871 Zbl 0189.31703

[Freiberg 2012] T. Freiberg, "Products of shifted primes simultaneously taking perfect power values", *J. Aust. Math. Soc.* **92**:2 (2012), 145–154. MR 2999152 Zbl 06124076

[Olson 1969] J. E. Olson, "A combinatorial problem on finite abelian groups, II", *J. Number Theory* **1** (1969), 195–199. MR 39 #1552 Zbl 0167.28004

Received: 2013-10-19 Revised: 2014-08-29 Accepted: 2014-09-07

elesluis@gmail.com

Departamento de Matemáticas Aplicadas, Universidad de Istmo, Ciudad Universitaria S/N, Barrio Santa Cruz, 4a Sección, Santo Domingo, 70110 Tehuantepec, Oaxaca, Mexico

florian.luca@wits.ac.za

School of Mathematics, University of the Witwatersrand, P.O. Box Wits 2050, Johannesburg, 2000 South Africa

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.749

On the cardinality of infinite symmetric groups

Matt Getzen

(Communicated by Kenneth S. Berenhaut)

A new proof is given that the symmetric group of any set X with three or more elements, finite or infinite, has cardinality strictly greater than that of X. Use of the axiom of choice is avoided throughout.

John Dawson and Paul Howard [1976] proved that the symmetric group of any set X with three or more elements, finite or infinite, has cardinality strictly greater than that of X. Significantly, their proof does not rely upon the axiom of choice. However, it does rely upon Cantor's theorem that the power set of any set X, finite or infinite, has cardinality strictly greater than that of X. We give a new proof of Dawson and Howard's result that relies upon neither the axiom of choice nor Cantor's theorem.

Recall that $\mathrm{Sym}(X)$ is the *symmetric group* of X, that is the set of all bijections between a set X and itself under function composition. More specifically, we call each bijection between a set and itself a *permutation*, each element that is mapped to itself by a permutation a *fixed point*, each pair of elements that are mapped to one another by a permutation a *transposition*, and each permutation that is its own inverse an *involution*.

The following results can easily be obtained and are listed without proof: (i) every fixed point in a permutation is also a fixed point in that permutation's inverse; (ii) every transposition in a permutation is also a transposition in that permutation's inverse; (iii) every permutation is an involution if and only if it is made up entirely of fixed points and transpositions; (iv) for all sets X, there exists an injection from X into $\mathrm{Sym}(X)$; and (v) in the case of all sets X with three or more elements, $\mathrm{Sym}(X)$ contains at least three involutions.

Theorem. *For any set X with three or more elements, finite or infinite,* $\mathrm{Sym}(X)$ *has cardinality strictly greater than that of X.*

Proof. We proceed by contradiction. Assume that there does exist a bijection $\mathcal{F}$ from X to $\mathrm{Sym}(X)$, and construct the permutation $\star$ in $\mathrm{Sym}(X)$ as follows:

MSC2010: primary 03E99; secondary 20B30, 03E25.
Keywords: set theory, infinite symmetric groups, axiom of choice.

(1) Let a, b, and c be three elements of X such that $\mathcal{F}(a)$, $\mathcal{F}(b)$, and $\mathcal{F}(c)$ are all involutions in $\mathrm{Sym}(X)$ with

$$\star(a) = b,$$
$$\star(b) = c,$$
$$\star(c) = a.$$

(2) For every other element i of X such that $\mathcal{F}(i)$ is an involution in $\mathrm{Sym}(X)$, but i is not equal to a, b, or c, we have

$$\star(i) = i.$$

(3) For each pair of permutations σ and μ in $\mathrm{Sym}(X)$ that are one another's inverses, and for each pair of elements s and m of X such that $\mathcal{F}(s) = \sigma$ and $\mathcal{F}(m) = \mu$, if σ transposes s and m then we have $\star(s) = s$ and $\star(m) = m$, but if σ does not transpose s and m then we have $\star(s) = m$ and $\star(m) = s$. In other words,

$$\star(s) = \begin{cases} s & \text{if } \sigma(s) = m \text{ and } \sigma(m) = s, \\ m & \text{if } \sigma(s) \neq m \text{ or } \sigma(m) \neq s, \end{cases}$$

$$\star(m) = \begin{cases} m & \text{if } \sigma(s) = m \text{ and } \sigma(m) = s, \\ s & \text{if } \sigma(s) \neq m \text{ or } \sigma(m) \neq s. \end{cases}$$

Note that $\star$ is a permutation of X and therefore an element in $\mathrm{Sym}(X)$. Note also that $\star$ is not an involution and therefore must have a distinct inverse, call it $\star^{-1}$. Thus, some element of X must be the preimage of $\star$ under $\mathcal{F}$. Let n denote just such an element of X. Additionally, some element of X other than n must be the preimage of $\star^{-1}$ under $\mathcal{F}$. Let w denote just such an element of X. That is, $\mathcal{F}(n) = \star$ and $\mathcal{F}(w) = \star^{-1}$. As $\star$ and $\star^{-1}$ are of the same general form as σ and μ above, it now follows that

$$\star(n) = \begin{cases} n & \text{if } \star(n) = w \text{ and } \star(w) = n, \\ w & \text{if } \star(n) \neq w \text{ or } \star(w) \neq n, \end{cases}$$

$$\star(w) = \begin{cases} w & \text{if } \star(n) = w \text{ and } \star(w) = n, \\ n & \text{if } \star(n) \neq w \text{ or } \star(w) \neq n. \end{cases}$$

In other words, assuming that the bijection $\mathcal{F}$ does in fact exist, n and w will be transposed with one another in $\star$ if and only if n and w are not transposed with one another in $\star$, a contradiction! Therefore no such bijection exists between X and $\mathrm{Sym}(X)$. Conversely, as we already know that there does exist an injection from X into $\mathrm{Sym}(X)$, we conclude that $\mathrm{Sym}(X)$ must have cardinality strictly greater than that of X. □

Through showing that the power set of any set X, finite or infinite, has cardinality strictly greater than that of X, Georg Cantor revolutionized mathematics and inspired the field of set theory. It is interesting to wonder how different the world might have been if mathematicians' first forays into the higher realms of the infinite had been inspired not by power sets, but by symmetric groups.

Acknowledgements

The author would like to thank John Dawson and Paul Howard for commenting on an earlier draft of this paper, and Rommy Marquez, Xizhong Zheng, Louis Friedler, Ned Wolff, and especially Carlos Ortiz for all their help and encouragement.

References

[Dawson and Howard 1976] J. W. Dawson, Jr. and P. E. Howard, "Factorials of infinite cardinals", *Fund. Math.* **93**:3 (1976), 185–195. MR 55 #7779 Zbl 0365.02050

Received: 2014-01-07 Accepted: 2014-07-12

mgetzen@arcadia.edu

Department of Mathematics & Computer Science,
Arcadia University, 450 South Easton Road,
Glenside, PA 19038, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.753

Adjacency matrices of zero-divisor graphs of integers modulo n

Matthew Young

(Communicated by Kenneth S. Berenhaut)

We study adjacency matrices of zero-divisor graphs of $\mathbb{Z}_n$ for various n. We find their determinant and rank for all n, develop a method for finding nonzero eigenvalues, and use it to find all eigenvalues for the case $n = p^3$, where p is a prime number. We also find upper and lower bounds for the largest eigenvalue for all n.

1. Introduction

Let R be a commutative ring with a unity. The notion of a *zero-divisor graph* of R was pioneered by Beck [1988]. It was later modified by Anderson and Livingston [1999] to be the following.

Definition 1.1. *The zero-divisor graph* $\Gamma(R)$ of the ring R is a graph with the set of vertices $V(R)$ being the set of zero-divisors of R and edges connecting two vertices $x, y \in R$ if and only if $x \cdot y = 0$.

To each (finite) graph Γ, one can associate the *adjacency matrix* $A(\Gamma)$ that is a square $|V(\Gamma)| \times |V(\Gamma)|$ matrix with entries $a_{ij} = 1$, if v_i is connected with v_j, and zero otherwise. In this paper we study the adjacency matrices of zero-divisor graphs $\Gamma_n = \Gamma(\mathbb{Z}_n)$ of rings $\mathbb{Z}_n$ of integers modulo n, where n is not prime. We note that the adjacency matrices of zero-divisor graphs of $\mathbb{Z}_p \times \mathbb{Z}_p$, $\mathbb{Z}_p[i]$, and $\mathbb{Z}_p[i] \times \mathbb{Z}_p[i]$, where p is a prime number and $i^2 = -1$, were studied in [Sharma et al. 2011].

2. Properties of adjacency matrices of Γ_n

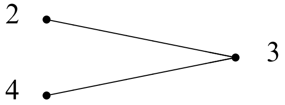
Let $n = p_1^{t_1} \cdots p_s^{t_s}$, where $p_1, \dots, p_s$ are distinct primes. For any divisor d of n, we define $S(d) = \{k \in \mathbb{Z}_n \mid \gcd(k, n) = d\}$. If $d = p_1^{a_1} \cdots p_s^{a_s}$, we will also write $S(d) = S(a_1, \dots, a_s)$. We can easily compute the size of the sets $S(d)$.

Proposition 2.1. *For a divisor d of n, the cardinality of the set $S(d)$ is equal to $|S(d)| = \phi(n/d)$, where ϕ denotes Euler's totient function.*

MSC2010: 05C50, 13M99.
Keywords: adjacency matrix, zero-divisor graph.

Proof. A positive integer m less than n is contained in the set $S(d)$ if and only if $\gcd(n,m)=d$, which happens if and only if $m=\hat{m}d$ and $\gcd(n/d,\hat{m})=1$. Thus, there is a one-to-one correspondence between the element m of $S(d)$ and integers $\hat{m}$, where $0<\hat{m}<n/d$ and $\gcd(n/d,\hat{m})=1$. □

Example 2.2. We illustrate Proposition 2.1 with the zero-divisor graph Γ_6 and its adjacency matrix:



$$A(\Gamma_6)=\begin{bmatrix}0&1&0\\1&0&1\\0&1&0\end{bmatrix},\quad S(2)=\{2,4\},\quad |S(2)|=\phi(6/2)=2.$$

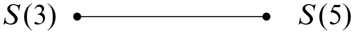
Theorem 2.3. *If* $n>4$, *then* $\det A(\Gamma_n)=0$.

Proof. Write $n=p_1^{t_1}p_2^{t_2}\cdots p_s^{t_s}$ as above. For $i=1,\dots,s$, if $p_i>2$, then $|S(n/p_i)|=p_i-1>1$. If a vertex v of Γ_n corresponding to some divisor d of n is adjacent to one of the elements of $S(n/p_i)$, then $d\cdot(n/p_i)=0 \pmod{n}$. Thus the product of d with any multiple of n/p_i is also zero, and v is adjacent to every element of $S(n/p_i)$. So $A(\Gamma_n)$ will have repeated rows corresponding to each element of $S(n/p_i)$. Since $|S(n/p_i)|>1$, we conclude that $\det A(\Gamma_n)=0$.

If $n{=}2^t$, we must have $t>2$. Then $6\in S(2)$, and $|S(2)|>1$. So $A(\Gamma_n)$ will have repeated rows corresponding to 2 and 6, and $\det A(\Gamma_n)=0$. □

The sets $S(d)$ for all divisors d of a given integer n are an *equitable partition* of the set of vertices $V(\Gamma_n)$. That is, any two vertices in $S(d_i)$ have the same number of neighbors in $S(d_j)$ for all divisors d_i,d_j of n. This allows us to define a *projection graph* $\pi\Gamma_n$ as a graph with vertices $S(d)$ for all $d|n$ and edges connecting $S(d_i)$ with $S(d_j)$ if every element in $S(d_i)$ is connected with every element in $S(d_j)$ in Γ_n.

Example 2.4. The projection graph $\pi\Gamma_{15}$:



Proposition 2.5. *The number of vertices in the graph* $\pi\Gamma_n$, *where* $n=\prod_{i=1}^{s}p_i^{t_1}$, *is*

$$|V(\pi\Gamma_n)|=\prod_{i=1}^{s}(t_i+1)-2.$$

Proof. The vertices of $\pi\Gamma_n$ are the sets $S(d)$ that are in one-to-one correspondence with the divisors d of n. If the prime decomposition for n is $n=\prod_{i=1}^{s}p_i^{t_i}$ and the prime decomposition for a divisor d of n is $d=\prod_{i=1}^{s}p_i^{a_i}$, then we have t_i+1

choices for the exponent a_i of p_i in d. The choice of all $a_i = 0$ leads to $d = 1$, and the choice of each $a_i = t_i$ leads to $d = n$, neither of which are proper divisors of n. So the number of proper divisors d is $\prod_{i=1}^{s}(t_i + 1) - 2$. □

Let $A(\pi(\Gamma_n))$ denote the adjacency matrix of $\pi(\Gamma_n)$. We will also consider the weighted adjacency matrix $\mathcal{A}(\pi(\Gamma_n))$, where $a_{ij} = |S(d_j)|$ whenever $S(d_i)$ is connected with $S(d_j)$. In the above example,

$$A(\pi(\Gamma_{15})) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathcal{A}(\pi(\Gamma_{15})) = \begin{bmatrix} 0 & 2 \\ 4 & 0 \end{bmatrix}.$$

The following theorem relates the ranks of the various adjacency matrices.

Theorem 2.6. *Let $n = \prod_{i=1}^{s} p_i^{t_i}$. Then,*

$$\operatorname{rank} A(\Gamma_n) = \operatorname{rank} A(\pi\Gamma_n) = \operatorname{rank} \mathcal{A}(\pi\Gamma_n) = \prod_{i=1}^{s}(t_i + 1) - 2.$$

Proof. Recall that vertices in $V(\pi\Gamma_n)$ correspond to sets $S(d)$, where $d|n$. Since each element of $S(d)$ contributes exactly the same row to the adjacency matrix $A(\Gamma_n)$, it follows that $\operatorname{rank} A(\Gamma_n) \leq |V(\pi\Gamma_n)|$. On the other hand, $\operatorname{rank} A(\pi\Gamma_n) \leq \operatorname{rank} A(\Gamma_n)$ since we just remove repeated rows and columns to get $A(\pi\Gamma_n)$ from $A(\Gamma_n)$. Obviously, $\operatorname{rank} A(\pi\Gamma_n) = \operatorname{rank} \mathcal{A}(\pi\Gamma_n)$. So it is enough to show $\operatorname{rank} A(\pi\Gamma_n) = |V(\pi\Gamma_n)|$.

Let $n = p_1^{t_1} p_2^{t_2} \cdots p_s^{t_s}$. Since the rank of the matrix does not change with permutations of rows, assume that the rows of $A(\pi\Gamma_n)$ correspond to the $S(d)$, with $d|n$, in the following order: $S(p_1)$, $S(n/p_1)$, $S(p_2)$, $S(n/p_2)$, $\ldots$, $S(p_s)$, $S(n/p_s)$, $S(p_i p_j)$ with i and j not necessarily distinct, $S(n/(p_i p_j))$ for all possible pairs of i and j, $S(p_i p_j p_k)$, $S(n/(p_i p_j p_k))$ for all possible triples i, j, k with i, j, k not necessarily distinct, etc.

We will compute the determinant of $A(\pi\Gamma_n)$ and, by showing that it is not zero, will prove that $|V(\pi\Gamma_n)|$ rows of $A(\pi\Gamma_n)$ are linearly independent.

The first row corresponding to $S(p_1)$ has 1 in the second column corresponding to $S(n/p_1)$ and the rest of the entries are 0. Expand the determinant of $A(\pi\Gamma_n)$ along the first row to get

$$\det A(\pi\Gamma_n) = -\det A_{1,2},$$

where $-\det A_{1,2}$ is the cofactor of the $(1, 2)$ entry of $A(\pi\Gamma_n)$. Note that the first column of $A_{1,2}$ has 1 in the first row and the rest of the entries are 0. Expand $\det A_{1,2}$ along the first column to get $\det A_{1,2} = \det A^{(2)}$, where $A^{(2)}$ is a matrix obtained from $A(\pi\Gamma_n)$ by deleting the first two rows and columns. So, we can conclude that $\det A(\pi\Gamma_n) = -\det A^{(2)}$.

We repeat this procedure for all $S(p_i)$ and $S(n/p_i)$ to get

$$\det A(\pi\Gamma_n) = (-1)^s \det A^{(2s)},$$

where $A^{(2s)}$ is obtained from $A(\pi\Gamma_n)$ by deleting the first $2s$ rows and columns.

Now consider $S(p_i p_j)$. In $\pi\Gamma_n$, the vertex corresponding to $S(p_i p_j)$ is adjacent to vertices of $S(n/p_i)$, $S(n/p_j)$ and $S(n/(p_i p_j))$. However, in the matrix $A^{(2s)}$, the row corresponding to $S(p_i p_j)$ will have only one 1 in the column corresponding to $S(n/(p_i p_j))$ since the columns corresponding to $S(n/p_i)$ and $S(n/p_j)$ were deleted. So we can repeat the procedure of expanding the determinant along the rows and columns corresponding to $S(p_i)$ to expand the determinant along the rows and then the columns corresponding to $S(n/(p_i p_j))$.

Then continue to $S(n/(p_i p_j p_k))$ in a similar fashion. In the end we will be left either with a 2×2 matrix of determinant -1, or a 1×1 matrix of determinant 1. So $\det A(\pi\Gamma_n) = (-1)^m$, where m is the number of distinct divisors d of n such that $d < \sqrt{n}$. It follows that $\operatorname{rank} A(\pi\Gamma_n) = |V(\pi\Gamma_n)|$. The result follows from Proposition 2.5. □

Corollary 2.7. *We have* $\det A(\pi\Gamma_n) = (-1)^m$, *where* $m = \lfloor(\operatorname{rank} A)/2\rfloor$.

Proof. This follows from the proof of Theorem 2.6. □

Corollary 2.8. *We have* $\det \mathcal{A}(\pi\Gamma_n) = (-1)^m \prod_{d|n} |S(d)|$, *where* $m = \lfloor(\operatorname{rank} \mathcal{A})/2\rfloor$.

Proof. This result follows from the previous corollary and the fact that $\mathcal{A}(\pi\Gamma_n)$ is obtained from $A(\pi\Gamma_n)$ by multiplying the j-th column by $|S(d_j)|$. □

Corollary 2.9. *The multiplicity of the eigenvalue* 0 *of* $A(\Gamma_n)$ *is*

$$n - \phi(n) - \prod_{i=1}^{s}(t_i + 1) + 2.$$

Proof. The multiplicity of the eigenvalue 0 of $A(\Gamma_n)$ is $|V(\Gamma_n)| - |V(\pi(\Gamma_n))|$. The number of vertices of Γ_n is the number of positive integers less than n and not relatively prime to n, which is $n - \phi(n)$. The number of vertices of $\pi\Gamma_n$ is

$$\prod_{i=1}^{s}(t_i + 1) - 2.$$ □

Part of the following result is known, but we will include it for the convenience of the reader.

Proposition 2.10. *A nonzero* $\lambda \in \mathbb{R}$ *is an eigenvalue of* $A(\Gamma_n)$ *if and only if it is an eigenvalue of* $\mathcal{A}(\pi(\Gamma_n))$.

Thus, to find all the nonzero eigenvalues of $A(\Gamma_n)$, it is enough to find all the eigenvalues of $\mathcal{A}(\pi(\Gamma_n))$. The proposition makes it especially easy when n has few factors. When $n = pq$ is a product of distinct primes, Γ_n is a bipartite graph. It is known (and easy to see) that the nonzero eigenvalues of $\mathcal{A}(\Gamma_{pq})$ are $\pm\sqrt{(p-1)(q-1)}$ with multiplicity 1. When $n = p^2$, the matrix $\mathcal{A}(\Gamma_n)$ is a 1×1 matrix with $p-1$ as a sole entry and hence the eigenvalue. So we consider the following two examples.

Example 2.11. Consider Γ_{p^3}. The matrix $\mathcal{A}(\Gamma_{p^3})$ takes the form

$$\mathcal{A} = \begin{bmatrix} 0 & p-1 \\ p(p-1) & p-1 \end{bmatrix}.$$

Its characteristic polynomial is $p(\lambda) = \lambda^2 - (p-1)\lambda - p(p-1)^2$, and the eigenvalues are $\lambda_{1,2} = \frac{1}{2}(p-1)(1 \pm \sqrt{1+4p})$.

Example 2.12. Consider Γ_{p^2q}, where p and q are distinct primes. In this case,

$$\mathcal{A}(\Gamma_{p^2q}) = \begin{bmatrix} 0 & p-1 & 0 & 0 \\ (p-1)(q-1) & p-1 & q-1 & 0 \\ 0 & p-1 & 0 & p^2-p \\ 0 & 0 & q-1 & 0 \end{bmatrix}.$$

The characteristic polynomial of this matrix is

$$p(\lambda) = \lambda^4 - (p-1)\lambda^3 - 2p(p-1)(q-1)\lambda^2 + p^2(p-1)(q-1)\lambda + p(p-1)^2(q-1)^3.$$

The roots can be found by the formulas for the roots of the fourth degree polynomial, but are too cumbersome to include here.

3. Estimates on eigenvalues

Since the increase in the number of factors of n leads to a rapid increase of the size of the adjacency matrix and the degree of the characteristic polynomial, one can use some known results to approximate the nonzero eigenvalues of $A(\Gamma_n)$. Since $A(\Gamma_n)$ is symmetric, all its eigenvalues are real numbers. We will number them from largest to smallest $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$.

The *degree* of a vertex of a graph is the number of the edges incident to this vertex. Given a (finite) graph Γ, let $\operatorname{maxdeg}(\Gamma) = \max\{\deg v \mid v \in V(\Gamma)\}$ and

$$\operatorname{avedeg}(\Gamma) = \frac{\sum_{v \in V(\Gamma)} \deg v}{|V(\Gamma)|}.$$

It is known (see [Brouwer and Haemers 2012, Proposition 3.1.2]) that

$$\operatorname{avedeg}(\Gamma) \leq \lambda_1 \leq \operatorname{maxdeg}(\Gamma).$$

We next compute $\operatorname{maxdeg}(\Gamma_n)$ and $\operatorname{avedeg}(\Gamma_n)$.

Proposition 3.1. *Let $n = p_1^{t_1} \cdots p_s^{t_s}$ with $p_1 < p_2 < \cdots < p_s$. Then*

$$\operatorname{maxdeg}(\Gamma_n) = n/p_1 - 1.$$

Proof. For a divisor d of n, denote the corresponding vertex in Γ_n by v_d. The vertex v_d is connected to a vertex v_c corresponding to a divisor c of n by an edge in Γ_n if and only if $dc \equiv 0 \pmod n$. This happens if and only if c is a multiple of n/d (and is less than n). There are $d-1$ such multiples, and so $\deg(v_d) = d-1$. Then the vertex with the largest degree will correspond to the largest divisor of n, which is n/p_1. Since for any $d|n$, the degrees of all vertices in $S(d)$ are the same and the sets $S(d)$ partition $V(\Gamma_n)$, $\operatorname{maxdeg}(\Gamma_n) = n/p_1 - 1$. □

Proposition 3.2. *The average degree of the graph Γ_n is*

$$\operatorname{avedeg}(\Gamma_n) = \frac{\sum_{d|n, d\neq n} \phi(n/d)(d-1)}{n - \phi(n) - 1}.$$

Proof. To compute the average degree, we take the degree $d-1$ of a vertex in $S(d)$, multiply by the cardinality $\phi(n/d)$ of the set $S(d)$, sum these products over all proper divisors d of n, and divide by the total number of vertices $n - \phi(n) - 1$. □

The estimate of the eigenvalue λ_1 using the average degree of the graph is inconvenient to use. So we use the results on interlacing and on bipartite subgraphs of Γ_n for alternative estimates that are easier to use. Let Γ be any graph and Δ an *induced subgraph*, that is, a subgraph obtained from Γ by deleting some vertices and all edges incident to the deleted vertices. The following result is known (see [Brouwer and Haemers 2012, Proposition 3.2.1] or [Godsil and Royle 2001, Theorem 9.1.1]). Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ be eigenvalues of $A(\Gamma)$ and $\theta_1 \geq \cdots \geq \theta_l$ be the eigenvalues of $A(\Delta)$; then $\lambda_i \geq \theta_i \geq \lambda_{k-l+i}$ for $i = 1, 2, \ldots, l$. With the interlacing result in mind, we prove the following proposition.

Proposition 3.3. *Let λ_1 be the largest eigenvalue of $\mathcal{A}(\Gamma_n)$:*

(1) *If n is a product containing two or more distinct primes, then $\lambda_1 \geq \sqrt{\phi(n)}$.*

(2) *If $n = p^t$, then $\lambda_1 \geq p^{\lceil t/2 \rceil} - 1$.*

Proof. Let $n = p_1^{t_1} p_2^{t_2} \cdots p_s^{t_s}$. Consider the bipartite subgraph of Γ_n induced by the vertices in the sets $S(p_1^{t_1})$ and $S(n/p_1^{t_1})$. The largest eigenvalue μ_1 of this subgraph is

$$\mu_1 = \sqrt{|S(p_1^{t_1})| \left| S\left(\frac{n}{p_1^{t_1}}\right) \right|} = \sqrt{\phi\left(\frac{n}{p_1^{t_1}}\right) \phi(p_1^{t_1})} = \sqrt{\phi(n)}.$$

The interlacing results give $\mu_1 \leq \lambda_1$.

To prove the second statement, notice that all vertices of $S(p^i)$ are connected to all the vertices of $S(p^j)$ whenever $i+j \geq t$, $1 \leq i, j \leq t-1$. In the case t is even, Γ_n contains a complete subgraph induced by vertices in the sets $S(p^{t/2}), \dots, S(p^{t-1})$. The largest eigenvalue μ_1 of this complete subgraph equals the number of vertices in the subgraph, which we compute next. By Proposition 2.1, $|S(d)| = \phi(n/d)$, so the number of vertices in the complete subgraph will be $\phi(p)+\phi(p^2)+\dots+\phi(p^{t/2})$. This can be expressed as $(p-1)+p(p-1)+p^2(p-1)+\dots+p^{t/2-1}(p-1)$, which after summing the arithmetic progression becomes $p^{t/2}-1=\mu_1$. In the case of t odd, Γ_n contains a complete subgraph induced by vertices in the sets $S(p^{\lceil t/2 \rceil}), \dots, S(p^{k-1})$. Using Proposition 2.1 and summing up the number of vertices as in case of t even gives $p^{\lfloor t/2 \rfloor} = \mu_1$. □

We can use the above estimates on the largest eigenvalue λ_1 of $A(\Gamma_n)$ to prove that there are only finitely many graphs with small eigenvalues.

Theorem 3.4. *For any positive integer k, there exists only a finite number of integers n such that all the eigenvalues of $A(\Gamma_n)$ are less or equal than k.*

Proof. We will use the estimates μ_1 on λ_1 obtained in Proposition 3.3. If, for a given $k>0$, we have $\lambda_1 \leq k$, then $\mu_1 \leq k$. We will show $\mu_1 \leq k$ is only possible for a finite number of integers n. Suppose $n = p^t$. Then $p^{t/2}-1$ must be less than or equal to k. Thus $t \leq 2\log_p(k+1)$, and there are only finitely many such positive integers.

Now suppose n is divisible by at least two distinct primes, say $n = p_1^{t_1} \cdots p_s^{t_s}$. The Euler function on n can be computed as

$$\phi(n) = \phi(p_1^{t_1}) \cdots \phi(p_s^{t_s}) = p_1^{t_1-1}(p_1-1) \cdots p_s^{t_s-1}(p_s-1).$$

Since there are only finitely many primes less than $k+1$, and only a finite number of possible exponents $t_1, \dots, t_s$ that satisfy $0 < t_1, \dots, t_s \leq 2\log_2 k$, there are only finitely many positive integers n such that $\phi(n) \leq k^2$. □

Example 3.5. We will find all positive integers n such that all the eigenvalues of $A(\Gamma_n)$ are less than or equal to $k=2$. Suppose $n=p^t$. If t is even, then we must have $p^{t/2}-1 \leq 2$. The only such possibilities are $p=2, t=2$ and $p=3, t=2$. If t is odd, we must have $p^{\lfloor t/2 \rfloor} \leq 2$. This happens only if $p=2, t=3$. If n is divisible by at least two distinct primes, we must have $\phi(n) \leq 4$, and computations show that this is satisfied only for $n = 6, 10, 12$. For $n = 4, 6, 8, 9, 10, 12$, we compute the eigenvalues of $A(\Gamma_n)$ and see that for all above n except $n=12$, the eigenvalues are less or equal than 2. Thus, the adjacency matrices of Γ_4, Γ_6, Γ_8, Γ_9 and Γ_{10} have all their eigenvalues less or equal than 2.

Example 3.6. For $k=3$, the graphs all of whose eigenvalues are less than or equal to 3, in addition to the graphs of Example 3.5, are Γ_{12}, Γ_{14} and Γ_{15}. For $k=4$, the new graphs (in addition to the ones with eigenvalues less or equal to 3) are Γ_{16}, Γ_{21}, Γ_{22}, Γ_{25}, Γ_{26} and Γ_{34}.

As we were considering the above examples, we noticed that for adjacency matrices of even rank, precisely half of the nonzero eigenvalues were positive and half negative. For adjacency matrices of odd rank, we always had one more positive eigenvalue than negative. We investigate this further.

The independence number $\alpha(\Gamma)$ of a graph Γ is the size of the largest set of pairwise nonadjacent vertices. Let r denote the number of eigenvalues of a (weighted) adjacency matrix $A(\Gamma)$ of a graph Γ, $r_+(A)$ the number of positive eigenvalues, and $r_-(A)$ the number of negative eigenvalues. It is known (see [Brouwer and Haemers 2012, Theorem 3.5.4]) that $\alpha(\Gamma) \le r - r_+(A)$ and $\alpha(\Gamma) \le r - r_-(A)$. We use this fact to show the following result.

Theorem 3.7. *Suppose the rank of $A(\Gamma_n)$ is r. Then $A(\Gamma_n)$ has $\lceil r/2 \rceil$ positive eigenvalues and $\lfloor r/2 \rfloor$ negative eigenvalues.*

Proof. We first note that it is enough to prove the theorem for $\mathcal{A}(\pi\Gamma_n)$ since $A(\Gamma_n)$ and $\mathcal{A}(\pi\Gamma_n)$ have the same nonzero eigenvalues. We start by computing the independence number of $\pi\Gamma_n$. Recall that the vertices of $\pi\Gamma_n$ are the sets $S(d)$, for all divisors d of n, and $S(d_1)$ is connected to $S(d_2)$ by an edge if n divides the product $d_1 d_2$. So for all $d \le \sqrt{n}$, the vertices $S(d)$ are pairwise not connected. It is easy to see that all the vertices of $\pi\Gamma_n$ can be split into pairs $S(d)$ and $S(n/d)$, with $S(\sqrt{n})$ without a pair if $\sqrt{n}$ is a divisor of n. So the set of $S(d)$ with $d|n$ and $d < \sqrt{n}$ is the maximal nonadjacent set of cardinality $\lfloor r/2 \rfloor$, where r denotes the number of vertices of $\pi\Gamma_n$. The number of vertices of $\pi\Gamma_n$ is equal to the rank of $\mathcal{A}(\pi\Gamma_n)$ by Theorem 2.6. So the independence number $\alpha(\pi\Gamma_n)$ is equal to $\lfloor r/2 \rfloor$.

For r even, we have $r/2 \le r - r_+(\mathcal{A})$ and $r/2 \le r - r_-(\mathcal{A})$, which implies the statement of the theorem. For r odd, it remains to show that we have one more positive eigenvalue than negative. By Corollary 2.8, we know that the sign of $\det \mathcal{A}(\pi\Gamma_n)$ is given by $(-1)^{\lfloor r/2 \rfloor}$. On the other hand, $\det \mathcal{A}(\pi\Gamma_n)$ is equal to the product of eigenvalues. So the parity of the number of negative eigenvalues must determine the sign of $(-1)^{\lfloor r/2 \rfloor}$. Since $\lfloor r/2 \rfloor \le r - r_-(\mathcal{A})$ and $\lfloor r/2 \rfloor \le r - r_+(\mathcal{A})$, we must have $r_-(\mathcal{A}) = \lfloor r/2 \rfloor$. □

Acknowledgement

I offer my utmost gratitude to Dr. Lucy Lifschitz who introduced this topic to me, taught me everything I know about it that I didn't discover for myself, asked all of the right questions that led me to my results, provided all of the sources and previously known results, and converted my disorganized original draft into a cohesive, legible paper. None of this would have been possible without her guidance.

References

[Anderson and Livingston 1999] D. F. Anderson and P. S. Livingston, "The zero-divisor graph of a commutative ring", *J. Algebra* **217**:2 (1999), 434–447. MR 2000e:13007 Zbl 0941.05062

[Beck 1988] I. Beck, "Coloring of commutative rings", *J. Algebra* **116**:1 (1988), 208–226. MR 89i:13006 Zbl 0654.13001

[Brouwer and Haemers 2012] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*, Springer, New York, 2012. MR 2882891 Zbl 1231.05001

[Godsil and Royle 2001] C. Godsil and G. Royle, *Algebraic graph theory*, Graduate Texts in Mathematics **207**, Springer, New York, 2001. MR 2002f:05002 Zbl 0968.05002

[Sharma et al. 2011] P. Sharma, A. Sharma, and R. K. Vats, "Analysis of adjacency matrix and neighborhood associated with zero divisor graph of finite commutative rings", *Int. J. Comput. Appl.* **14**:3 (2011), Article #7.

Received: 2014-02-07 Revised: 2014-08-27 Accepted: 2014-09-19

mjy5068@psu.edu

Mathematics Department, Penn State University,
University Park, 008 McAllister Building,
State College, PA 16802, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.763

Expected maximum vertex valence in pairs of polygonal triangulations

Timothy Chu and Sean Cleary

(Communicated by Kenneth S. Berenhaut)

Edge-flip distance between triangulations of polygons is equivalent to rotation distance between rooted binary trees. Both distances measure the extent of similarity of configurations. There are no known polynomial-time algorithms for computing edge-flip distance. The best known exact universal upper bounds on rotation distance arise from measuring the maximum total valence of a vertex in the corresponding triangulation pair obtained by a duality construction. Here we describe some properties of the distribution of maximum vertex valences of pairs of triangulations related to such upper bounds.

1. Introduction

Binary trees are widely used in a broad spectrum of computational settings. Binary search trees underlie many modern structures devoted to efficient searching, for example. Shapes of binary trees affect the performance of searches, and there have been a wide variety of approaches to ensure such efficiency. Natural dual structures to rooted ordered binary trees are triangulations of polygons with a marked edge or vertex. Rotations in binary trees correspond to edge-flip moves in such triangulations of polygons, so the rotation distance between two rooted ordered binary trees corresponds exactly to the edge-flip distance between the two corresponding triangulations of marked polygons.

Properties of rotations have been widely studied; see Knuth [1973] for background and fundamental algorithms. There is no known polynomial-time algorithm for computing rotation (or equivalently, edge-flip) distance, though there are a variety of efficient approximation algorithms [Baril and Pallo 2006; Cleary and St. John 2010; 2009]. A straightforward argument of Culik and Wood [1982] shows that for

MSC2010: 05C05, 68P05.

Keywords: random binary trees.

The authors gratefully acknowledge support from the National Science Foundation under Grant Nos. #0811002 and #1417820. This work was partially supported by a grant from the Simons Foundation (#234548 to Sean Cleary).

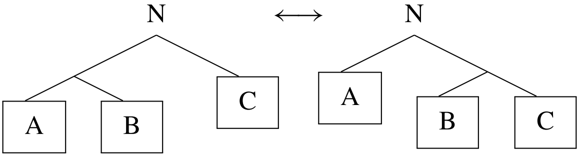


Figure 1. Rotation at a node N. Right rotation at N transforms the left tree to the right one, and left rotation at N is the inverse operation which transforms the right tree to the left one. A, B, and C represent leaves or subtrees, and the node N could be at the root or any other position in the tree.

any two trees with n internal nodes, there is always a path of length at most $2n-2$. Sleator, Tarjan and Thurston [Sleator et al. 1988] showed that the distance is never more than $2n-6$ using an argument described below based upon maximum summed vertex valence in the pair of triangulations, and furthermore that for all very large n, that bound is achieved. Recently, Pournin [2014] showed that, in fact, the upper bound is achieved for all $n \geq 11$.

A rotation move in a rooted binary tree relative to a fixed node N is a promotion of one grandchild node of N to a child node of N, a demotion of a child of N to a grandchild of N, and a switch of parent node for one grandchild of N, preserving order. This occurs in the vicinity of a single node, as pictured in Figure 1. The corresponding edge-flip move in a triangulation occurs in a single quadrilateral formed by two triangles which share an edge. The common edge between two adjacent triangles is exchanged for the opposite diagonal in that quadrilateral, as shown in Figure 2. If the edge-flip distance between two triangulations of a regular polygon is k, that means that there is a sequence of k edge flips transforming the first triangulation to the second and there is no shorter sequence accomplishing the same transformation.

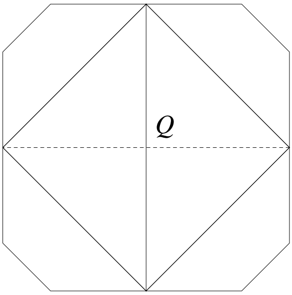


Figure 2. An edge flip across a quadrilateral Q. The four peripheral quadrilaterals denote (possibly empty) triangulated polygons whose triangulations are unchanged by the edge flip in quadrilateral Q.

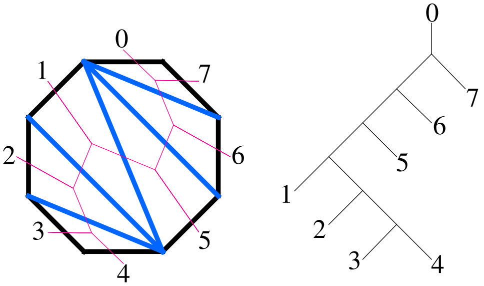


Figure 3. A triangulation of the octagon and the corresponding dual tree, with sides numbered to match the leaves. Pulling up on the edge from the marked side of the octagon (marked as leaf 0) gives the tree on the right.

2. Triangular subdivisions of polygons

Here, by a *triangulation of size n*, we mean a triangulation of $n-1$ interior edges subdividing a regular (n+2)-gon, where we choose to label vertices from 0 to $n+1$. Such a triangulation is dual to a tree with $n+1$ leaves and n internal nodes, with leaves labeled from 0 to n. See Figure 3.

The number of triangulations of size n is the n-th Catalan number, C_n, and since C_n grows exponentially at rate of $4^n n^{-3/2}$, the number of pairs of trees of size n grows on the order of $16^n n^{-3}$. Because of the rapid growth of the number of tree pairs (or equivalently, triangulation pairs), computing these quantities exhaustively via complete enumeration is not feasible beyond small n. To explore this exponentially growing space, we use sampling techniques to characterize the expected behavior of randomly selected triangulation pairs. We experiment computationally by choosing pairs of triangulations of size n uniformly at random, computing the relevant vertex sums and tabulating the results. As in [Chu and Cleary 2013], we use the linear-time random tree-generation procedure of Rémy [1985] to generate efficiently ordered trees uniformly at random, rather than considering the Yule distribution on tree pairs studied by Cleary, Passaro and Toruno [Cleary et al. 2015]. The quantities studied here are the maximum valence sums of vertices, described in the next section.

3. Vertex valence sums

Vertex valence sums play a role in the upper bounds for edge-flip distance. Given a pair of triangulations S and T, we count the number of interior edges s_i and t_i incident to each vertex i in the polygon and form the *vertex valence sum* for vertex i as $s_i + t_i$. See Figure 4. The bound of $2n-6$ for $n \geq 11$ from [Sleator et al. 1988] is obtained by the following argument. There are a total of $n-1$ edges in each triangulation, each with 2 endpoints, giving $4n-4$ total endpoints of interior edges

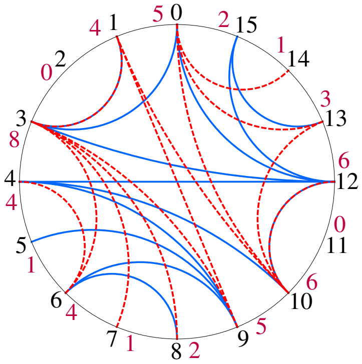


Figure 4. Two superimposed triangulations, one drawn in dashed red and one in solid blue. The vertices are numbered and the total valences are given in purple for each vertex. For example, the total valence of vertex 0 is five, with 3 red edges and 2 blue edges incident there. In this example, the summed vertex valences range from a low of 0 (at vertex 2, indicating a common peripheral triangle) to a high of 8, which occurs at vertex 3.

for the pair of triangulations. Each endpoint occurs at one of the $n+2$ vertices, so the average valence of a vertex is merely $(4n-4)/(n+2) = 4-12/(n+2)$. For $n \geq 11$, this gives average valence larger than 3 (approaching 4 in the limit of large n). Since the summed valences can only be integers, if the sum is more than 3, there must be a vertex j of total valence 4 or more. We consider a path of triangulations from S to T by way of a fan triangulation F_j, which is the fan from vertex j (that is, a triangulation of the polygon where every triangle has an edge incident on vertex j). Transforming S to F_j takes exactly $n-1-s_j$ edge flips, as

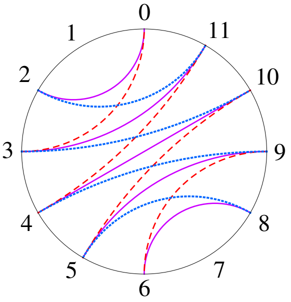


Figure 5. Two superimposed zigzag triangulations, one drawn in dotted blue and one in dashed red, with common segments in solid purple. Each red dashed edge can be directly flipped to the corresponding blue one, giving an edge-flip distance of 4 between the two triangulations, despite the vertex valence sums commonly being 4.

there is always an edge flip which increases the valence of vertex j by one and there are exactly that many edges to flip. Similarly, from T to F_j there is a path of length $n-1-t_j$ and such a path is minimal. So there is a path from S to F_j to T of length no more than $2n-2-s_j-t_j$, giving the $2n-6$ bound in the case that the maximum vertex valence s_j+t_j is exactly 4. In cases where the maximal vertex valence is higher, the upper bound is correspondingly decreased.

We note that for a triangulation pair, the vertex sums need to be quite evenly distributed around the polygon to have a chance of being maximally distant for that size. The average valence sum is between 3 and 4 for $n \geq 11$ and no maximally distant pair of triangulations can have any vertex sums of 5 or larger. Note that having a maximal vertex sum of 4 is necessary but not sufficient for being a maximally distant triangulation pair, as can be easily seen by considering a zigzag triangulation of a regular n-gon beginning at vertex 0 and a reflected zigzag triangulation beginning at vertex 2, as shown in Figure 5. The maximal vertex sum is 4 but the edge-flip distance between these two is less than $n/2$ (flipping the red edges to the blue edges in the figure), far less than that of the maximum possible. There are many other configurations with maximal vertex 4 which are not remotely close to the $2n-6$ upper bound as well. Nevertheless, if there is even a single vertex with summed valence 5 or more, the two triangulations cannot be at the maximal $2n-6$ distance.

4. Discussion

There is an obviously increasing relationship between triangulation size and expected maximum observed summed valence across the vertices. However, the growth rate appears slow, as the relationship appears to be either straightening or slightly convex downward in log-scaled Figure 6, where a straight line would indicate logarithmic growth. The experimental evidence suggests that the relationship is at most logarithmic. Figure 7 shows an example of a distribution of maximum summed valence for a particular size, $n = 950$, showing the shape of typical distributions that arise in these computations.

A *combing triangulation with respect to* v is a triangulation where every edge is incident with the vertex v. Using the argument of [Sleator et al. 1988], an upper bound on rotation distance from S to T comes from the path which first flips successively edges in S to be incident with v, a vertex of maximum summed valence, to obtain the combing triangulation for v. Then the path goes from that combing triangulation to T, successively flipping to edges in T which are not incident on v. At each step of the resulting path, there is at least one edge in S which can be flipped to be incident to v or one edge incident to v which can be flipped to an edge in T. The resulting length of the path is the number of edges in S and T which are not incident on v. Combinatorial arguments give that there is always a vertex of

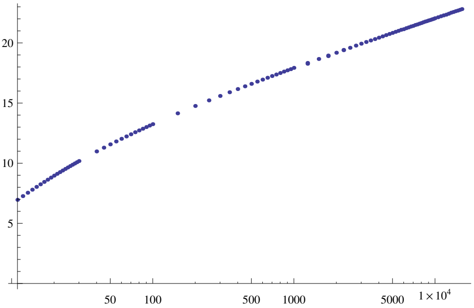


Figure 6. Maximum valence increases with triangulation size. Here we show average maximum summed vertex valence vertically, against the size of triangulations plotted horizontally on a logarithmic scale. For each size, the number of triangulation pairs sample to estimate the average maximum vertex valence ranges from 100,000 to 10 million depending upon size.

summed degree 4, giving the universal (for $n \geq 11$) bound of $2n-6$. For larger summed valence k for a particular pair of trees, the same arguments show that an upper bound for distance for that pair is $2n-k-2$. The experimental data in Table 1 shows that though it is common for randomly selected tree pairs to have higher summed valence than the minimum of 4, it is often not markedly higher than 4. In the case of randomly selected tree pairs, we know from the asymptotic analysis of Cleary, Rechnitzer and Wong [Cleary et al. 2013] that for large n, two randomly selected triangulations of size n are likely to have about $(16/\pi - 5)n \cong 0.093n$ common edges. Thus, the upper bounds for rotation distance arising from common edges, edges which are a single rotation from a common edge, and from common components of small size are generally much stronger than those upper bounds arising from the path through a fan on a vertex of maximum summed valence.

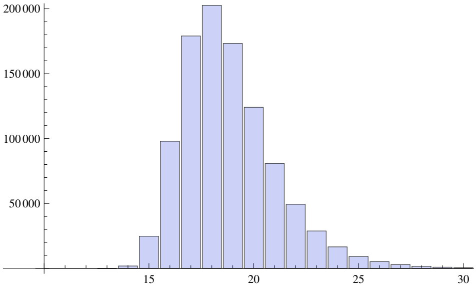


Figure 7. An example of the distribution of maximum vertex valence for one million triangulation pairs, for size 950, with average 17.8 and standard deviation 2.26.

triangulation size	average max vertex sum	σ max vertex sum
15	8.03203	1.55756
20	8.96568	1.70762
30	10.1791	1.87928
40	10.9805	1.97628
50	11.5674	2.03815
75	12.5759	2.12531
100	13.2461	2.16791
200	14.7589	2.23289
500	16.5983	2.26295
1000	17.9277	2.27465
2000	19.184	2.24797
3000	19.9369	2.2556
5000	20.8352	2.24551
7500	21.5392	2.23559
10000	22.0527	2.2108
12000	22.3505	2.20521
15000	22.7512	2.20811

Table 1. Observed averages and standard deviations of maximum vertex sums via experiments involving 10 million ($n \leq 30$), 1 million ($n < 1000$) or 100,000 (for $n \geq 10000$) runs depending upon the triangulation sizes.

n	fraction with max vertex sum 4
11	0.0050032
12	0.0015352
13	0.0004462
14	0.0001232
15	0.000035
16	$8.1 \cdot 10^{-6}$
17	$2.4 \cdot 10^{-6}$
18	$5.0 \cdot 10^{-7}$
19	$1.0 \cdot 10^{-7}$
20+	none observed

Table 2. Fractions of triangulations with maximum vertex sum exactly 4. Hundreds of millions were considered for $n \geq 20$, finding none selected at random.

The work of Pournin [2014] constructs carefully very specific examples of triangulations which are at maximal distance $2n-6$ for all $n \geq 11$. One question is how common such maximally distant pairs are. The experimental evidence in Table 2 shows that examples with this extremal behavior are quite rare. In the language of associahedra, used in [Pournin 2014], pairs of triangulations at maximal $2n-6$ distance correspond to antipodal points, which have many long geodesics between them, with typically many of those passing through the distinguished "fan" triangulations. But from the analysis here, it appears quite rare that a pair of triangulations will have a geodesic which passes through a fan point, indicating further the rarity of these extremal antipodal pairs.

References

[Baril and Pallo 2006] J.-L. Baril and J.-M. Pallo, "Efficient lower and upper bounds of the diagonal-flip distance between triangulations", *Inform. Process. Lett.* **100**:4 (2006), 131–136. MR 2007e:68070 Zbl 1185.68845

[Chu and Cleary 2013] T. Chu and S. Cleary, "Expected conflicts in pairs of rooted binary trees", *Involve* **6**:3 (2013), 323–332. MR 3101764 Zbl 1274.05066

[Cleary and St. John 2009] S. Cleary and K. St. John, "Rotation distance is fixed-parameter tractable", *Inform. Process. Lett.* **109**:16 (2009), 918–922. MR 2010g:68114 Zbl 1205.68531

[Cleary and St. John 2010] S. Cleary and K. St. John, "A linear-time approximation for rotation distance", *J. Graph Algorithms Appl.* **14**:2 (2010), 385–390. MR 2011m:68205 Zbl 1215.68271

[Cleary et al. 2013] S. Cleary, A. Rechnitzer, and T. Wong, "Common edges in rooted trees and polygonal triangulations", *Electron. J. Combin.* **20**:1 (2013), Paper 39. MR 3035049 Zbl 1267.05249

[Cleary et al. 2015] S. Cleary, J. Passaro, and Y. Toruno, "Average reductions between random tree pairs", *Involve* **8**:1 (2015), 63–69. MR 3321710 Zbl 1309.05158

[Culik and Wood 1982] K. Culik, II and D. Wood, "A note on some tree similarity measures", *Inform. Process. Lett.* **15**:1 (1982), 39–42. MR 83k:68059 Zbl 0489.68058

[Knuth 1973] D. E. Knuth, *The art of computer programming, 3: Sorting and searching*, Addison-Wesley, Reading, MA, 1973. 2nd ed. in 1997. MR 56 #4281 Zbl 0302.68010

[Pournin 2014] L. Pournin, "The diameter of associahedra", *Adv. Math.* **259** (2014), 13–42. MR 3197650 Zbl 1292.52011

[Rémy 1985] J.-L. Rémy, "Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire", *RAIRO Inform. Théor.* **19**:2 (1985), 179–195. MR 87h:68132 Zbl 0565.05037

[Sleator et al. 1988] D. D. Sleator, R. E. Tarjan, and W. P. Thurston, "Rotation distance, triangulations, and hyperbolic geometry", *J. Amer. Math. Soc.* **1**:3 (1988), 647–681. MR 90h:68026 Zbl 0653.51017

Received: 2014-03-06 Revised: 2014-10-17 Accepted: 2014-11-22

timchu100@gmail.com
Department of Computer Science,
The City College of New York, City University of New York,
New York, NY 10031, United States

cleary@sci.ccny.cuny.edu
Department of Mathematics,
The City College of New York and the CUNY Graduate Center,
City University of New York, NAC R8133,
160 Convent Avenue, New York, NY 10031, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.771

Generalizations of Pappus' centroid theorem via Stokes' theorem

Cole Adams, Stephen Lovett and Matthew McMillan

(Communicated by Kenneth S. Berenhaut)

This paper provides a novel proof of a generalization of Pappus' centroid theorem on n-dimensional tubes using Stokes' theorem on manifolds.

1. Introduction

The (second) Pappus centroid theorem or the Pappus–Guldin theorem states that the volume of a solid of revolution generated by rotating a plane region $\mathcal{R}$ with piecewise-smooth boundary about an axis L is $2\pi r$ Area($\mathcal{R}$), where r is the distance from the centroid of $\mathcal{R}$ to L. This result generalizes considerably to the following main theorem.

Theorem 1.1 (main theorem). *Let C be a simple, regular, smooth curve in $\mathbb{R}^n$. Let $\mathcal{R}$ be a region in $\mathbb{R}^{n-1}$ whose boundary is an embedding of the $(n-2)$-dimensional sphere $\mathbb{S}^{n-2}$. Let $\mathcal{W}$ be a region in $\mathbb{R}^n$ whose boundary is a generalized tube around C such that the cross-section normal to C of $\mathcal{W}$ at each point P of C is the region $\mathcal{R}$ with centroid at P. Assuming the cross-section $\mathcal{R}$ rotates smoothly as it "travels" along C, then*

$$\mathrm{Vol}_n(\mathcal{W}) = \mathrm{length}(C)\,\mathrm{Vol}_{n-1}(\mathcal{R}).$$

The Pappus centroid theorem follows from this main theorem by taking $n = 3$, C to be a circle in $\mathbb{R}^3$, and $\mathcal{R}$ to remain fixed with respect to the principal normal to C in the normal plane. This theorem recently was proved by Gray, Miquel, and Domingo-Juan in [Domingo-Juan and Miquel 2004] and [Gray and Miquel 2000] using parallel transport. However, Goodman and Goodman [1969] proved this theorem in a special case for $\mathbb{R}^3$ using elementary methods related to Stokes' theorem. This article proves the main theorem in full generality using Stokes' theorem on manifolds. In this

MSC2010: 53A07, 58C35.

Keywords: Stokes' theorem on manifolds, volume, manifolds, tubes.

The authors thank Wheaton College's Faculty Student Mentoring Initiative, which made possible the collaboration that led to this article.

regard, we can consider the proof elementary compared to those in [Domingo-Juan and Miquel 2004] and [Gray and Miquel 2000].

Before proving the main theorem in full generality, we sketch the proof of it in $\mathbb{R}^3$ found in [Goodman and Goodman 1969], leaving the reader to consult that work for details. The description of the generalized tube and the method involving the divergence theorem motivate the situation for arbitrary n.

2. Generalized tubes in dimension 3

Definition 2.1. Let C be a simple, regular, smooth space curve and let $\mathcal{R}$ be a compact planar region with one boundary component $\partial\mathcal{R}$, a piecewise smooth simple closed curve. Select a marked point P in $\mathcal{R}$. C has a normal plane at each point. Let $\mathcal{W}$ be a region in $\mathbb{R}^3$ such that the intersection of $\mathcal{W}$ with the normal plane to C at any point is isometric to the region $\mathcal{R}$, with the corresponding marked point P lying on the curve C. We assume $\mathcal{R}$ rotates smoothly in the normal plane to C as it travels along C. Such a region $\mathcal{W}$ is called a *generalized tube* along C with cross-section $\mathcal{R}$ and center P.

This definition allows for rotational freedom of $\mathcal{R}$ around the marked point P in the normal planes to C. However, this rotational varies smoothly. We may also describe the generalized tube as a fiber-bundle over C with fiber $\mathcal{R}$, that is a subbundle of the normal bundle over C.

Figure 1 depicts two generalized tubes around a portion of a helix. More precisely, the figure depicts the tube boundary excluding the "caps", or cross-sections at the end points of C. The planar curve shows its generating region $\mathcal{R}$ where the marked point of $\mathcal{R}$ is the origin.

Let $\mathcal{S}$ be the boundary $\partial\mathcal{W}$ of a generalized tube excluding the caps. (If C is a closed curve, then $\partial\mathcal{W}$ has no caps.) Suppose that $\alpha : [0,\ell] \to \mathbb{R}^3$ gives a parametrization by arclength of C. Also suppose that $\vec{\beta} : [0,c] \to \mathbb{R}^2$ is a parametrization of $\partial\mathcal{R}$ placing the marked point P at the origin. We write $\vec{\beta}(u) = (x(u), y(u))$ for the coordinate functions. A parametrization for $\mathcal{S}$ is

$$\begin{aligned}\vec{X}(s,u) = \vec{\alpha}(s) &+ \big(\cos(\theta(s))x(u) - \sin(\theta(s))y(u)\big)\vec{P}(s)\\ &+ \big(\sin(\theta(s))x(u) + \cos(\theta(s))y(u)\big)\vec{B}(s) \qquad (1)\end{aligned}$$

for some function $\theta(s)$, where $\vec{P}(s)$ and $\vec{B}(s)$ are respectively the principal normal and binormal vector functions to $\vec{\alpha}(s)$.

Recall that $(\vec{T}(s), \vec{P}(s), \vec{B}(s))$, where $\vec{T}$, $\vec{P}$, and $\vec{B}$ are the usual tangent, principal normal, and binormal vectors to $\vec{\alpha}(s)$, is called the Frenet frame to $\vec{\alpha}(s)$. The function $\theta(s)$ determines the rotation of the region $\mathcal{R}$ around the origin with respect to the Frenet frame. The stipulation that $\mathcal{R}$ rotates smoothly as it moves along C implies that $\theta(s)$ is a smooth function.

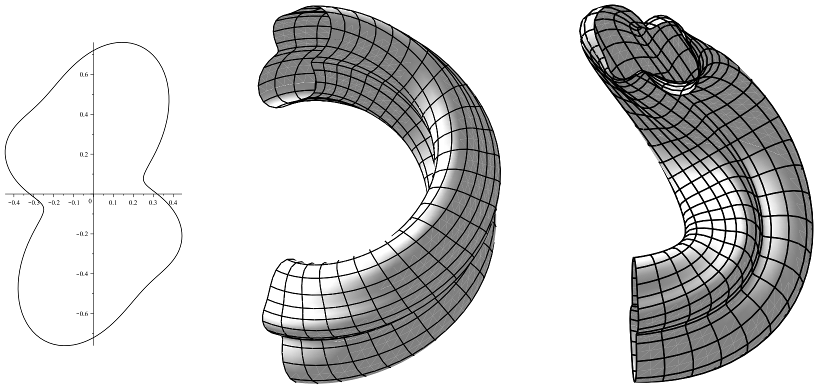


Figure 1. A generalized tube with its generating region.

Figure 1, middle, depicts a generalized tube where the x-axis in the depiction of $\mathcal{R}$ always lies along the principal normal vector of $\vec{\alpha}(s)$, and Figure 1, right, depicts a generalized tube with the same cross-section region but having some rotation with respect to the basis $(\vec{P}, \vec{B})$ in the normal plane. For brevity, we write

$$\vec{X} = \vec{\alpha} + (x\cos\theta - y\sin\theta)\vec{P} + (x\sin\theta + y\cos\theta)\vec{B},$$

where functional dependence is understood from (1).

Theorem 2.2 [Goodman and Goodman 1969, Corollary 2]. *The volume of a generalized tube as described in Definition 2.1 is* $V = \operatorname{length}(C)\operatorname{Area}(\mathcal{R})$.

The Goodmans’ method to calculate the volume uses the fact that the position vector field $\vec{r}(x, y, z) = (x, y, z)$ has divergence equal to 3 everywhere. So, using the notation defined above, the volume of the generalized tube is

$$\operatorname{Vol}(\mathcal{W}) = \frac{1}{3}\iiint_{\mathcal{W}} 3\,dV = \frac{1}{3}\iiint_{\mathcal{W}} \nabla\cdot\vec{r}\,dV = \frac{1}{3}\iint_{\partial\mathcal{W}} \vec{r}\cdot d\vec{A},$$

where $d\vec{A}$ is the outward pointing surface element. Note that $\partial\mathcal{W}$ consists of the tube’s outward surface $\mathcal{S}$, parametrized by $\vec{X}$, and the end caps (if C is not a closed curve). Over $\mathcal{S}$, $d\vec{A}$ is given by $d\vec{A} = (\vec{X}_u \times \vec{X}_s)\,du\,ds$ with $(u, v) \in [0, c]\times[0, \ell]$, while on the end caps, $d\vec{A} = -\vec{T}(0)\,dA$ when $s = 0$ and $d\vec{A} = \vec{T}(\ell)\,dA$ when $s = \ell$. The caps, like any cross-section at s, are parametrized by

$$\vec{Y}_s(p, q) = \vec{\alpha}(s) + p\vec{P}(s) + q\vec{B}(s) \quad \text{for } (p, q) \in \mathcal{R}_s,$$

where $\mathcal{R}_s$ is the region $\mathcal{R}$ rotated about the origin (the marked point P) by the angle $\theta(s)$. Thus, since $\vec{T}(s)$ is perpendicular to both $\vec{P}(s)$ and $\vec{B}(s)$, we have

$$\begin{aligned}
&3\,\mathrm{Vol}(\mathcal{W})\\
&\quad=\int\limits_{s=0}^{\ell}\int\limits_{u=0}^{c}\vec{X}\cdot(\vec{X}_u\times\vec{X}_s)\,du\,ds+\iint\limits_{\mathcal{R}_\ell}\vec{\alpha}(\ell)\cdot\vec{T}(\ell)\,dp\,dq+\iint\limits_{\mathcal{R}_0}-\vec{\alpha}(0)\cdot\vec{T}(0)\,dp\,dq\\
&\quad=\int\limits_{s=0}^{\ell}\int\limits_{u=0}^{c}\vec{X}\cdot(\vec{X}_u\times\vec{X}_s)\,du\,ds+\mathrm{Area}(\mathcal{R})\big(\vec{\alpha}(\ell)\cdot\vec{T}(\ell)-\vec{\alpha}(0)\cdot\vec{T}(0)\big). \qquad (2)
\end{aligned}$$

The problem of calculating the volume of $\mathcal{W}$ reduces to calculating the double integral in (2).

Recall that vectors of the Frenet frame (parametrized by arclength) differentiate according to

$$\begin{aligned}
\vec{T}' &= \phantom{-\kappa\vec{T}} \quad \kappa\vec{P},\\
\vec{P}' &= -\kappa\vec{T} \qquad\qquad +\tau\vec{B},\\
\vec{B}' &= \phantom{-\kappa\vec{T}} \quad -\tau\vec{P},
\end{aligned} \qquad (3)$$

where $\kappa(s)$ and $\tau(s)$ are the curvature and torsion functions of the space curve $\vec{\alpha}(s)$. Then the tangent vectors to $\vec{X}$ are given (after simplification) by

$$\begin{aligned}
\vec{X}_u &= (x'\cos\theta-y'\sin\theta)\vec{P}+(x'\sin\theta+y'\cos\theta)\vec{B},\\
\vec{X}_s &= (1-\kappa x\cos\theta+\kappa y\sin\theta)\vec{T}-(\theta'+\tau)(x\sin\theta+y\cos\theta)\vec{P}\\
&\qquad+(\theta'+\tau)(x\cos\theta-y\sin\theta)\vec{B}.
\end{aligned}$$

So

$$\begin{aligned}
\vec{X}_u\times\vec{X}_s=(\theta'+\tau)(xx'+yy')\vec{T}+(1-\kappa x\cos\theta+\kappa y\sin\theta)(x'\sin\theta+y'\cos\theta)\vec{P}\\
-(1-\kappa x\cos\theta+\kappa y\sin\theta)(x'\cos\theta-y'\sin\theta)\vec{B}.
\end{aligned}$$

The dot product $\vec{X}\cdot(\vec{X}_u\times\vec{X}_s)$ involves many terms. However, all of the additive terms involved in the integrals are multiplicatively separable, which, by the usual corollary to Fubini's theorem, allows us to separate the double integral. Many of the integrals involving u vanish or evaluate to a simple constant, namely the area of the cross-section. Consider the following integrals. By substitution,

$$\int_{u=0}^{c} xx'\,du = x^2\Big|_0^c = 0$$

because $(x(u), y(u))$ with $u\in[0,c]$ parametrizes a closed curve $\partial\mathcal{R}$. By similar reasoning, the following integrals are all 0:

$$\int_{u=0}^{c} x'\,du=0,\quad \int_{u=0}^{c} y'\,du=0,\quad \int_{u=0}^{c} xx'\,du=0,\quad \int_{u=0}^{c} yy'\,du=0. \qquad (4)$$

By Green's theorem for the area of the interior of a simple closed piecewise smooth curve,

$$\int_{u=0}^{c} xy'\,du = -\int_{u=0}^{c} yx'\,du = \iint_{\mathcal{R}} 1\,dA = \text{Area}(\mathcal{R}). \tag{5}$$

Also by Green's theorem,

$$\int_{u=0}^{c} \tfrac{1}{2}x^2y'\,du = \int_{u=0}^{c} -xyx'\,du = \iint_{\mathcal{R}} x\,dA = 0 \tag{6}$$

because this integral is the y-moment of $\mathcal{R}_s$ and by hypothesis, the centroid of $\mathcal{R}_s$ is $(0,0)$ for all s. By the same reasoning but for the x-moment, we also have

$$\int_{u=0}^{c} -\tfrac{1}{2}y^2x'\,du = \int_{u=0}^{c} xyy'\,du = \iint_{\mathcal{R}} y\,dA = 0. \tag{7}$$

Upon applying these integrals, only a few terms remain in (2). Setting $A = \text{Area}(\mathcal{R})$, we get

$$3\,\text{Vol}(\mathcal{W}) = \int_{s=0}^{\ell} (-\vec{\alpha}\cdot\vec{T}'A + 2A)\,ds + A\big(\vec{\alpha}(\ell)\cdot\vec{T}(\ell) - \vec{\alpha}(0)\cdot\vec{T}(0)\big).$$

Using integration by parts on the dot product, we obtain

$$\begin{aligned}
3\,\text{Vol}(\mathcal{W})
&= -A(\vec{\alpha}\cdot\vec{T})\big|_0^\ell + A\int_{s=0}^{\ell} \vec{\alpha}'\cdot\vec{T}\,ds + 2A\ell + A\big(\vec{\alpha}(\ell)\cdot\vec{T}(\ell) - \vec{\alpha}(0)\cdot\vec{T}(0)\big)\\
&= -A\big(\vec{\alpha}(\ell)\cdot\vec{T}(\ell) - \vec{\alpha}(0)\cdot\vec{T}(0)\big)\\
&\qquad + A\int_{s=0}^{\ell} \vec{T}\cdot\vec{T}\,ds + 2A\ell + A\big(\vec{\alpha}(\ell)\cdot\vec{T}(\ell) - \vec{\alpha}(0)\cdot\vec{T}(0)\big)\\
&= A\ell + 2A\ell = 3A\ell.
\end{aligned}$$

We conclude that $\text{Vol}(\mathcal{W}) = \text{Area}(\mathcal{R})\,\text{length}(C)$.

Theorem 2.2 establishes the main theorem of the paper for generalized tubes in $\mathbb{R}^3$. In order to prove the main theorem in full generality, we will need to use differential forms along with Stokes' theorem on manifolds. However, a key component to the main theorem is a set of integral formulas for the general case similar to (4), (5), (6), and (7).

3. Volumes, moments, and zero integrals for solids in $\mathbb{R}^m$

Recall that Stokes' theorem on manifolds states that if M is an m-dimensional, oriented manifold with boundary ∂M, and ω is a differential $(m-1)$-form on M, then

$$\int_{\partial M} \omega = \int_M d\omega, \tag{8}$$

where ∂M has the boundary orientation inherited from the orientation on M.

Definition 3.1. We define a *solid* in $\mathbb{R}^m$ as a compact embedded m-dimensional submanifold of $\mathbb{R}^m$ with boundary ∂M. We assume the pull-back orientation on M.

We define the $(m-1)$-form η^i in $\mathbb{R}^m$ by

$$\eta^i = (-1)^{i+1} dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m,$$

where $(y^1, y^2, \dots, y^m)$ is a coordinate system on $\mathbb{R}^m$ and $\widehat{\ }$ denotes removal of that term.

Lemma 3.2. *The m-dimensional volume of a solid M is*

$$\mathrm{Vol}_m(M) = \int_{\partial M} y^i \eta^i$$

for any $i = 1, 2, \dots, m$.

Proof. The differential of $y^i \eta^i$ is

$$d(y^i \eta^i) = (-1)^{i+1} dy^i \wedge dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m = dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m.$$

This form is precisely the volume form on $\mathbb{R}^m$, and thus on the solid M as well. Hence, by Stokes' theorem,

$$\int_{\partial M} y^i \eta^i = \int_M dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m = \mathrm{Vol}_m(M). \qquad \square$$

This lemma immediately implies the following corollary:

Corollary 3.3. *Let $\nu = \dfrac{1}{m} \sum_{i=1}^{m} y^i \eta^i$. The m-dimensional volume of M is*

$$\mathrm{Vol}_m(M) = \int_{\partial M} \nu.$$

In this article, if $F : M \to N$ is a differentiable map between differentiable manifolds, we will denote by $[dF]$ the matrix of functions of the differential dF in reference to given coordinate systems on M and on N. Furthermore, when the dimension of M is one less than the dimension of N and when coordinate systems on neighborhoods of M and N are implied, we denote by $|d_j F|$ the determinant of $[dF]$ in which the j-th row is removed.

Proposition 3.4. *Let M be an m-dimensional solid such that the boundary ∂M is the embedding of a continuous map $H : \mathbb{S}^{m-1} \to \mathbb{R}^m$ that is smooth except on a subset of measure 0 in $\mathbb{S}^{m-1}$. Suppose also that H induces an orientation on ∂M that is compatible with the boundary orientation induced from M. Let ν be the $(m-1)$-form as in Corollary 3.3 and let ω be the $(m-1)$-form on $\mathbb{S}^{m-1}$*

given by $\omega = dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{m-1}$ *for coordinates* $(x^1, x^2, \ldots, x^{m-1})$. *The* m*-dimensional volume of* M *is*

$$\mathrm{Vol}_m(M) = \int_{H(\mathbb{S}^{m-1})} \nu = \int_{\mathbb{S}^{m-1}} H^* \nu = \frac{1}{m} \int_{\mathbb{S}^{m-1}} \det(H, [dH])\, \omega, \tag{9}$$

where in $\det(H, [dH])$ *we write the components of* H *as a column vector. If* H *induces the opposite orientation, the second two integrals change sign.*

Proof. The equality

$$\mathrm{Vol}_m(M) = \int_{H(\mathbb{S}^{m-1})} \nu$$

follows immediately from Corollary 3.3. Let $(x^1, x^2, \ldots, x^{m-1})$ be coordinates on $\mathbb{S}^{m-1}$ and $(y^1, y^2, \ldots, y^m)$ on $\mathbb{R}^m$. Notice that the pullback of ν by H is

$$H^* \nu = \frac{1}{m} \sum_{i=1}^{m} H^i (-1)^{i+1} dH^1 \wedge dH^2 \wedge \cdots \wedge \widehat{dH^i} \wedge \cdots \wedge dH^m.$$

Or, writing in x^i coordinates, and using the fact that

$$dH^i = \frac{\partial H^i}{\partial x^j} dx^j$$

(assuming the Einstein summation convention), we find

$$H^* \nu = \frac{1}{m} \sum_{i=1}^{m} H^i (-1)^{i+1} \left(\frac{\partial H^1}{\partial x^{j_1}} dx^{j_1} \right) \wedge \left(\frac{\partial H^2}{\partial x^{j_2}} dx^{j_2} \right) \wedge \cdots \\ \wedge \widehat{\left(\frac{\partial H^i}{\partial x^{j_i}} dx^{j_i} \right)} \wedge \cdots \wedge \left(\frac{\partial H^m}{\partial x^{j_m}} dx^{j_m} \right).$$

By Theorem C.5.22 in [Lovett 2010], this is equivalent to

$$H^* \nu = \frac{1}{m} \sum_{i=1}^{m} H^i (-1)^{i+1} \begin{vmatrix} \frac{\partial H^1}{\partial x^1} & \frac{\partial H^1}{\partial x^2} & \cdots & \frac{\partial H^1}{\partial x^{m-1}} \\ \frac{\partial H^2}{\partial x^1} & \frac{\partial H^2}{\partial x^2} & \cdots & \frac{\partial H^2}{\partial x^{m-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\frac{\partial H^i}{\partial x^1}} & \widehat{\frac{\partial H^i}{\partial x^2}} & \cdots & \widehat{\frac{\partial H^i}{\partial x^{m-1}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H^m}{\partial x^1} & \frac{\partial H^m}{\partial x^2} & \cdots & \frac{\partial H^m}{\partial x^{m-1}} \end{vmatrix} dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{m-1}.$$

Taking the summation and recognizing the Laplace expansion of a determinant down the first column, we see that

$$H^*\nu = \frac{1}{m}\det(H,[dH])\,dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{m-1}.$$

Then (9) follows. Note that the second integral changes sign if H induces the opposite orientation on ∂M, so the third integral changes sign as well. □

Lemma 3.2, Corollary 3.3, and Proposition 3.4 are generalizations to higher dimensions of Green's theorem for area. For example, suppose that $\mathcal{S}$ is a solid in $\mathbb{R}^3$ such that the boundary $\partial\mathcal{S}$ is parametrized by $\vec{X}(u,v) = (x(u,v), y(u,v), z(u,v))$ with $(u,v) \in \mathcal{D}$ such that $\vec{X}_u \times \vec{X}_v$ is outward-pointing. Then by Proposition 3.4, the volume of $\mathcal{S}$ is

$$\mathrm{Vol}(\mathcal{S}) = \frac{1}{3}\iint_{\mathcal{D}} \begin{vmatrix} x & x_u & x_v \\ y & y_u & y_v \\ z & z_u & z_v \end{vmatrix}\, du\, dv.$$

Because of the flexibility in Stokes' theorem, as in Green's area theorem, this formula still applies when $\partial\mathcal{S}$ is piecewise smooth. In that case, we interpret the above integral as a sum of integrals taken over domains $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_r$ such that the parametrizations for the smooth pieces of $\partial\mathcal{S}$ have domains $\mathcal{D}_i$. The same principle applies in (9).

We will encounter other integrals that cancel. We list them here.

Proposition 3.5. *Let M be a solid and let $(y^1, y^2, \ldots, y^m)$ be a coordinate system on M. Then for i and q in $\{1, 2, \ldots, m\}$,*

$$\int_{\partial M} y^q \eta^i = \delta_i^q \,\mathrm{Vol}_m(M),$$

where δ_i^q is the Dirac delta in which $\delta_i^q = 1$ if $i = q$ and $\delta_i^q = 0$ if $i \neq q$.

Proof. The case with $i = q$ is Lemma 3.2. If $i \neq q$, then

$$d(y^q\eta^i) = dy^q \wedge dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m = 0$$

because one differential is repeated. Then by Stokes' theorem, we have

$$\int_{\partial M} y^q\eta^i = \int_M d(y^q\eta^i) = \int_M 0 = 0.$$ □

Corollary 3.6. *Let M, H, and ω be as in Proposition 3.4. Then*

$$\int_{\mathbb{S}^{m-1}} (-1)^{i+1} H^q |d_i H| \omega = \delta_i^q \,\mathrm{Vol}_m(M).$$

Proof. This follows immediately from the fact that

$$(-1)^{i+1} H^q |d_i H| \omega = H^*(y^q\eta^i).$$ □

Proposition 3.7. *Let M, H, and ω be as in Proposition 3.4. Let $\vec{a} = (a^1, a^2, \dots, a^m)$ be a constant vector, listed as a column vector. Then*

$$\int_{\mathbb{S}^{m-1}} \det(\vec{a}, [dH])\omega = 0.$$

Proof. By the reasoning in the proof of (9), we see that

$$\begin{aligned}\det(\vec{a}, [dH])\omega &= \sum_{i=1}^{m} (-1)^{i+1} a^i \, dH^1 \wedge dH^2 \wedge \cdots \wedge \widehat{dH^i} \wedge \cdots \wedge dH^m \\ &= H^* \Bigl(\sum_{i=1}^{m} (-1)^{i+1} a^i \, dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m \Bigr).\end{aligned}$$

Hence, by a pull-back and then Stokes' theorem,

$$\begin{aligned}\int_{\mathbb{S}^{m-1}} \det(\vec{a}, [dH]) &= \int_{H(\mathbb{S}^{m-1})} \sum_{i=1}^{m} (-1)^{i+1} a^i \, dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m \\ &= \int_M d\Bigl(\sum_{i=1}^{m} (-1)^{i+1} a^i \, dy^1 \wedge dy^2 \wedge \cdots \wedge \widehat{dy^i} \wedge \cdots \wedge dy^m \Bigr) \\ &= \int_M 0 \ = \ 0.\end{aligned}$$

□

In the proof of Theorem 2.2, certain integrals vanished by virtue of the cross-section always having its centroid on the curve C, and the same thing occurs in higher dimensions. The following proposition establishes the centroid generalizations needed later:

Proposition 3.8. *Let M be an m-dimensional solid as given in Definition 3.1. Let $(y^1, y^2, \dots, y^m)$ be a coordinate system covering M. Let $(\bar{y}^1, \bar{y}^2, \dots, \bar{y}^m)$ be the center of mass of M. Then*

$$\int_{\partial M} y^p y^q \eta^i = \begin{cases} 0 & \text{if } p \neq i \text{ and } q \neq i, \\ \bar{y}^p \operatorname{Vol}_m(M) & \text{if } p \neq i \text{ and } q = i, \\ 2\bar{y}^i \operatorname{Vol}_m(M) & \text{if } p = q = i. \end{cases}$$

Proof. By Stokes' theorem,

$$\int_{\partial M} y^p y^q \eta^i = \int_M d(y^p y^q \eta^i).$$

However,

$$d(y^p y^q \eta^i) = (y^q dy^p + y^p dy^q) \wedge \eta^i = y^q dy^p \wedge \eta^i + y^p dy^q \wedge \eta^i.$$

If neither $p=i$ nor $q=i$, then $dy^p \wedge \eta^i = 0$ and $dy^q \wedge \eta^i = 0$. If $q=i$ and $p \neq i$, then $d(y^p y^q \eta^i) = y^p \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m$ and

$$\int_M y^p \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m = \bar{y}^p \operatorname{Vol}_m(M)$$

by definition of the center of mass. Finally, if $p = q = i$, then $d(y^p y^q \eta^i) = 2y^i \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m$ and

$$\int_M 2y^i \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^m = 2\bar{y}^i \operatorname{Vol}_m(M). \qquad \square$$

4. Generalized tubes in higher dimensions

We are almost ready to prove Theorem 1.1. We must first set up a useful description of a generalized tube. Let $\mathcal{W}$ be a generalized tube with guiding curve C and cross-section $\mathcal{R}$ as described in the statement of the main theorem. A generalized tube is a fiber-bundle over C with fiber $\mathcal{R}$, that is, a subbundle of the normal bundle over C. Suppose that C is parametrized by arclength by $\alpha : [0, \ell] \to \mathbb{R}^n$. Suppose that the cross-section $\mathcal{R}$ is a solid in $\mathbb{R}^{n-1}$ whose boundary $\partial\mathcal{R}$ is parametrized by an orientation-preserving, differentiable map $H : \mathbb{S}^{n-2} \to \mathbb{R}^{n-1}$. We also assume that $\mathcal{R}$ rotates smoothly about the origin in the normal plane as it is transported along C. For the purpose of the theorem, we also assume that the center of mass of $\mathcal{R}$ is the origin in $\mathbb{R}^{n-1}$. Define $\overline{H} : \mathbb{S}^{n-2} \to \mathbb{R}^n$ by $\overline{H}(\vec{x}) = (0, H(\vec{x}))$.

The boundary $\partial\mathcal{W}$ of the solid generalized tube consists of the caps at $\alpha(0)$ and $\alpha(\ell)$ as well as the side surface $\mathcal{S}$, which we can parametrize by

$$\alpha(t) + M(t)\overline{H}(\vec{x}) \quad \text{for } (t, \vec{x}) \in [0, \ell] \times \mathbb{S}^{n-2},$$

where $M : [0, \ell] \to \mathrm{SO}(n)$ is a differentiable curve of special orthogonal (rotation) matrices in $\mathbb{R}^n$ such that for all t,

$$M(t) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = M(t)\vec{e}_1 = \alpha'(t). \tag{10}$$

Note that since $M(t)$ is a rotation matrix and the unit vector $\vec{e}_1$ in the y^1 direction is perpendicular to $\{(0, y^2, \ldots, y^n) \mid y^i \in \mathbb{R}\}$, then for all $t \in [0, \ell]$, the boundary of the cross-section $M(t)\overline{H}(\vec{x})$, for $\vec{x} \in \mathbb{S}^{n-2}$, is in a plane perpendicular to the tangent vector $\alpha'(t)$. For simplicity later, we write $F(t, \vec{x}) = M(t)\overline{H}(\vec{x})$.

Recall that since $M(t)$ is a special orthogonal matrix for all t, then $M(t)^{-1} = M(t)^\top$, $\det M(t) = 1$, and $M'(t) = M(t)A(t)$, where $A(t)$ is some antisymmetric matrix for all t. Using the rotation matrix $M(t)$ provides the following useful fact.

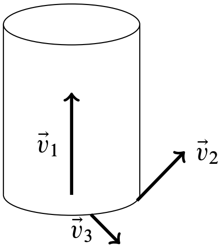


Figure 2. Reversed orientation on a cylinder.

Lemma 4.1. *The first component of the vector $M(t)^{-1}\alpha(t)$ is equal to the dot product $\alpha(t)\cdot\alpha'(t)$.*

Proof. By (10), the dot product $\alpha(t)\cdot\alpha'(t)$ is

$$\alpha(t)\cdot\alpha'(t)=\alpha(t)^\top M(t)\vec{e}_1.$$

Taking the transpose of the matrix expression on the right, and since the whole expression is just a real number, we get

$$\alpha(t)\cdot\alpha'(t)=\vec{e}_1^{\,\top}M(t)^\top\alpha(t)=(1\ \ 0\ \ \cdots\ \ 0)M(t)^{-1}\alpha(t),$$

and the lemma follows. □

Proof of Theorem 1.1. **Case 1:** Assume that the guiding curve C is not closed. Let ν be the $(n-1)$-form $\nu=\frac{1}{n}\sum_{i=1}^n y^i\eta^i$. By Corollary 3.3, the volume of the generalized tube is

$$\mathrm{Vol}_n(\mathcal{W})=\int_{\partial\mathcal{W}}\nu=\int_{\mathcal{S}}\nu+\int_{\mathrm{cap}_{t=0}}\nu+\int_{\mathrm{cap}_{t=\ell}}\nu. \tag{11}$$

We parametrize $\mathcal{S}$ by $\alpha+F$ but we note that this parametrization is orientation-reversing. This can be seen by applying our setup to the case of a circular cylinder in $\mathbb{R}^3$ and generalizing to higher dimensions. In Figure 2, $\vec{v}_1$ is $\alpha'(t)$, $\vec{v}_2$ is a tangent vector to the cross-section boundary in positive orientation, and $\vec{v}_3$ is the outward pointing normal vector to the solid M. These three vectors form a left-handed system so the orientation induced from our parametrization is reversed from the boundary orientation on ∂M induced by the standard orientation of $\mathbb{R}^n$ on M.

We can parametrize the caps by G_0 and G_ℓ where, for each $t\in[0,\ell]$, we define $G_t:\mathcal{R}\to\mathbb{R}^n$ with

$$G_t(\vec{z})=\alpha(t)+M(t)\begin{pmatrix}0\\ \vec{z}\end{pmatrix}.$$

Now G_0 induces an orientation that is opposite the boundary orientation on $\partial\mathcal{W}$, while G_t gives a compatible orientation. Hence, (11) becomes

$$\mathrm{Vol}_n(\mathcal{W})=-\int_{I\times\mathbb{S}^{n-2}}(\alpha+F)^*\nu-\int_{\mathcal{R}}G_0^*\nu+\int_{\mathcal{R}}G_\ell^*\nu. \tag{12}$$

We calculate the integrals on the caps first. By the same reasoning as in Proposition 3.4, for each $t \in [0, \ell]$,

$$G_t^* \nu = \frac{1}{n} \det(G_t, [d(G_t)])\, dz^1 \wedge dz^2 \wedge \cdots \wedge dz^{n-1}.$$

Now

$$\begin{aligned}\det(G_t, [d(G_t)]) &= \det\left(\alpha(t) + M(t)\begin{pmatrix}0\\ \vec{z}\end{pmatrix}, M(t)\begin{pmatrix}\vec{0}^\top\\ \boldsymbol{I}_{n-1}\end{pmatrix}\right)\\ &= \det(M(t)) \det\left(M(t)^{-1}\alpha(t) + \begin{pmatrix}0\\ \vec{z}\end{pmatrix}, \begin{pmatrix}\vec{0}^\top\\ \boldsymbol{I}_{n-1}\end{pmatrix}\right)\\ &= \det\left(M(t)^{-1}\alpha(t), \begin{pmatrix}0\\ \vec{z}\end{pmatrix}, \begin{pmatrix}\vec{0}^\top\\ \boldsymbol{I}_{n-1}\end{pmatrix}\right) = \alpha(t)\cdot\alpha'(t),\end{aligned}$$

where $\vec{0}^\top = (0, \ldots, 0)$, $\boldsymbol{I}_{n-1}$ is the $(n-1)\times(n-1)$ identity matrix and the last equality holds by Lemma 4.1. Consequently,

$$\int_{\mathcal{R}} G_\ell^* \nu - \int_{\mathcal{R}} G_0^* \nu = \frac{1}{n} \operatorname{Vol}_{n-1}(\mathcal{R})\big(\alpha(\ell)\cdot\alpha'(\ell) - \alpha(0)\cdot\alpha'(0)\big). \tag{13}$$

Now we must calculate $\int_{I\times\mathbb{S}^{n-2}} (\alpha + F)^*\nu$. Applying Proposition 3.4, over a coordinate patch of $\mathbb{S}^{n-2}$ with coordinate system $(x^1, x^2, \ldots, x^{n-2})$, we have

$$(\alpha+F)^*\nu = \frac{1}{n} \det\big(\alpha(t)+F(\vec{x}), \alpha'(t)+F_t(t,\vec{x}), M(t)[d\overline{H}]\big)\, dt\wedge dx^1\wedge\cdots\wedge dx^{n-2},$$

where here $F_t = \partial F/\partial t$. This can be broken down by multilinearity of the determinant as follows:

$$\begin{aligned}(\alpha+F)^*\nu = \frac{1}{n}\Big(&\det\big(\alpha(t), \alpha'(t), M(t)[d\overline{H}]\big)\\ &+\det\big(F(\vec{x}), F_t(t,\vec{x}), M(t)[d\overline{H}]\big) + \det\big(F(\vec{x}), \alpha'(t), M(t)[d\overline{H}]\big)\\ &+\det\big(\alpha(t), F_t(t,\vec{x}), M(t)[d\overline{H}]\big)\Big)\, dt\wedge dx^1\wedge\cdots\wedge dx^{n-2}.\end{aligned} \tag{14}$$

We now consider the integration over $[0, \ell]\times\mathbb{S}^{n-2}$ of the four forms in (14).

For the first determinant in (14),

$$\begin{aligned}\det\big(\alpha(t), \alpha'(t), M(t)[d\overline{H}]\big) &= \det\big(\alpha(t), M(t)\vec{e}_1, M(t)[d\overline{H}]\big)\\ &= \det(M(t))\det\big(M(t)^{-1}\alpha(t), \vec{e}_1, [d\overline{H}]\big)\\ &= -\det\big(\vec{e}_1, M(t)^{-1}\alpha(t), [d\overline{H}]\big).\end{aligned}$$

Doing Laplace expansion down the first column, we obtain an integral of the form in Proposition 3.7, with a vector $\vec{a}$ that depends on t. Hence, by Proposition 3.7,

$$\begin{aligned}&\int_{\mathbb{S}^{n-2}}\int_{t=0}^{\ell}\det\big(\alpha(t),\alpha'(t),M(t)[d\overline{H}]\big)\,dt\wedge dx^1\wedge\cdots\wedge dx^{n-2}\\ &\quad=(-1)^{n-2}\int_{t=0}^{\ell}\int_{\mathbb{S}^{n-2}}\det\big(\alpha(t),\alpha'(t),M(t)[d\overline{H}]\big)\,dx^1\wedge\cdots\wedge dx^{n-2}\wedge dt=0.\end{aligned}$$

For the second determinant in (14),

$$\begin{aligned}\det\big(F(t,\vec{x}),F_t(t,\vec{x}),M(t)[d\overline{H}]\big)&=\det\big(M(t)\overline{H}(\vec{x}),M(t)A(t)\overline{H}(\vec{x}),M(t)[d\overline{H}]\big)\\&=\det(M(t))\det\big(\overline{H}(\vec{x}),A(t)\overline{H}(\vec{x}),[d\overline{H}]\big)\\&=\det\big(\overline{H}(\vec{x}),A(t)\overline{H}(\vec{x}),[d\overline{H}]\big).\end{aligned}$$

Performing a Laplace expansion of the determinant using the first two columns of this last determinant produces terms similar to the forms described in Proposition 3.8. Since the centroid of $\mathcal{R}$ is assumed to be at the origin, then for all t, integrating all these terms over $\mathbb{S}^{n-2}$ gives 0.

For the third determinant in (14), we have

$$\begin{aligned}\det\big(F(t,\vec{x}),\alpha'(t),M(t)[d\overline{H}]\big)&=\det\big(M(t)\overline{H}(\vec{x}),M(t)\vec{e}_1,M(t)[d\overline{H}]\big)\\&=\det(M(t))\det\big(\overline{H}(\vec{x}),\vec{e}_1,[d\overline{H}]\big)\\&=-\det\big(\vec{e}_1,\overline{H}(\vec{x}),[d\overline{H}]\big)\\&=-\det(H(\vec{x}),[dH]),\end{aligned}$$

where the last equality follows by Laplace expansion of the determinant on the first row $(1\ \ 0\ \ \cdots\ \ 0)$. By Proposition 3.4,

$$\begin{aligned}&\int_{\mathbb{S}^{n-2}}\int_{t=0}^{\ell}\det\big(F(\vec{x}),\alpha'(t),[d(F)]\big)\,dt\wedge dx^1\wedge\cdots\wedge dx^{n-2}\\&\qquad=-\ell\int_{\mathbb{S}^{n-2}}\det\big(H(\vec{x}),[d(H)]\big)\,dx^1\wedge\cdots\wedge dx^{n-2}\\&\qquad=-(n-1)\ell\,\mathrm{Vol}_{n-1}(\mathcal{R}).\end{aligned}$$

As with previous determinants, the fourth determinant becomes

$$\det\big(\alpha(t),F_t(t,\vec{x}),M(t)[d\overline{H}]\big)=\det\big(M(t)^{-1}\alpha(t),A(t)\overline{H}(\vec{x}),[d\overline{H}]\big).$$

Since there are zeros in the first rows of $\overline{H}$ and $d(\overline{H})$, and because $M^{-1}=M^{\top}$, another Laplace expansion gives

$$\begin{aligned}\det\bigl(M^{-1}\alpha, A\overline{H}, [d\overline{H}]\bigr)&\\ &= \alpha_i M^i_1 \sum_{j=2}^{n} (-1)^j A^j_q \overline{H}^q |d_j \overline{H}| - \alpha_i \sum_{j=2}^{n} (-1)^j M^i_j A^1_q \overline{H}^q |d_j \overline{H}|\\ &= \sum_{j=2}^{n} \alpha_i \overline{H}^q |d_j \overline{H}| (-1)^j (M^i_1 A^j_q - M^i_j A^1_q),\end{aligned}$$

where we use the Einstein summation convention over the repeated indices appearing in superscript and subscript, namely i and q. By Proposition 3.5, after integration on $\mathbb{S}^{n-2}$, all terms will reduce to 0 except those for which $q=j$, which will give the volume of the cross-section, $\mathrm{Vol}_{n-1}(\mathcal{R})$. Set $I=[0,\ell]$. So,

$$\begin{aligned}\int_{\mathbb{S}^{n-2}} \int_I \det\bigl(M^{-1}\alpha, A\overline{H}, [d(\overline{H})]\bigr)\, dt\wedge dx^1 \wedge\cdots\wedge dx^{n-2}&\\ &= \int_{\mathbb{S}^{n-2}} \int_I \sum_{j=2}^{n} \alpha_i \overline{H}^q |d_j \overline{H}| (-1)^j (M^i_1 A^j_q - M^i_j A^1_q)\, dt\wedge dx^1 \wedge\cdots\wedge dx^{n-2}\\ &= \int_{\mathbb{S}^{n-2}} \int_I \sum_{j=2}^{n} \alpha_i \overline{H}^j |d_j \overline{H}| (-1)^j (M^i_1 A^j_j - M^i_j A^1_j)\, dt\wedge dx^1 \wedge\cdots\wedge dx^{n-2}\\ &= \int_{\mathbb{S}^{n-2}} \int_I \sum_{j=2}^{n} \alpha_i H^{j-1} |d_j \overline{H}| (-1)^j M^i_j A^j_1\, dt\wedge dx^1 \wedge\cdots\wedge dx^{n-2}\\ &= \int_I \alpha_i \sum_{j=2}^{n} M^i_j A^j_1\, dt \int_{\mathbb{S}^{n-2}} (-1)^j H^{j-1} |d_{j-1} H|\, dx^1 \wedge\cdots\wedge dx^{n-2}.\end{aligned}$$

But $\alpha'(t)=M(t)\vec{e}_1$, so $\alpha''(t)=M'(t)\vec{e}_1=M(t)A(t)\vec{e}_1$. Hence $M^i_j A^j_1$ are the components of the covariant vector $\alpha''(t)^{\top}$. Note that since $A(t)$ is an antisymmetric matrix, $A^1_1=0$. Thus $\alpha_i \sum_{j=2}^{n} M^i_j A^j_1 = \alpha(t)\cdot\alpha''(t)$. Hence, we get

$$\begin{aligned}\int_{\mathbb{S}^{n-2}} \int_I \det\bigl(M^{-1}\alpha, A\overline{H}, [d(\overline{H})]\bigr)\, dt\wedge dx^1 \wedge\cdots\wedge dx^{n-2}&\\ &= \mathrm{Vol}_{n-1}(\mathcal{R}) \int_I \alpha_i \alpha''_i\, dt\\ &= \mathrm{Vol}_{n-1}(\mathcal{R}) \biggl(\alpha(t)\cdot\alpha'(t)\Big|_0^{\ell} - \int_I \alpha'\cdot\alpha'\, dt \biggr)\\ &= \mathrm{Vol}_{n-1}(\mathcal{R}) \bigl(\alpha(t)\cdot\alpha'(t)\big|_0^{\ell} - \ell \bigr).\end{aligned}$$

Now putting into (12) the integrals of the four determinants in (14) and the integrals for the caps (13), we get

$$n\,\mathrm{Vol}_n(\mathcal{W}) = (n-1)\,\mathrm{Vol}_{n-1}(\mathcal{R})\ell - \mathrm{Vol}_{n-1}(\mathcal{R})\bigl(\alpha(t)\cdot\alpha'(t)\big|_0^\ell - \ell\bigr) \\ + \mathrm{Vol}_{n-1}(\mathcal{R})\bigl(\alpha(t)\cdot\alpha'(t)\big|_0^\ell\bigr). \quad (15)$$

Hence

$$\mathrm{Vol}_n(\mathcal{W}) = \mathrm{Vol}_{n-1}(\mathcal{R})\ell,$$

which establishes the main theorem when the guiding curve of the generalized tube is not closed.

Case 2: If C is a closed curve, then in (12) we do not have integrals for the caps. Then (15) becomes

$$n\,\mathrm{Vol}_n(\mathcal{W}) = (n-1)\,\mathrm{Vol}_{n-1}(\mathcal{R})\ell - \mathrm{Vol}_{n-1}(\mathcal{R})\bigl(\alpha(t)\cdot\alpha'(t)\big|_0^\ell - \ell\bigr),$$

and since $\alpha(0)\cdot\alpha'(0) = \alpha(\ell)\cdot\alpha'(\ell)$, the result of the main theorem follows for this case as well. □

References

[Domingo-Juan and Miquel 2004] M. C. Domingo-Juan and V. Miquel, "Pappus type theorems for motions along a submanifold", *Differential Geom. Appl.* **21**:2 (2004), 229–251. MR 2005e:53041 Zbl 1061.53038

[Goodman and Goodman 1969] A. W. Goodman and G. Goodman, "Generalizations of the theorems of Pappus", *Amer. Math. Monthly* **76** (1969), 355–366. MR 39 #2047 Zbl 0172.45902

[Gray and Miquel 2000] A. Gray and V. Miquel, "On Pappus-type theorems on the volume in space forms", *Ann. Global Anal. Geom.* **18**:3-4 (2000), 241–254. MR 2001i:53046 Zbl 1009.53027

[Lovett 2010] S. Lovett, *Differential geometry of manifolds*, A K Peters, Natick, MA, 2010. MR 2011k:53001 Zbl 1205.53001

Received: 2014-03-18 Accepted: 2014-11-22

zerg164@gmail.com
School of Engineering, Vanderbilt University,
Nashville, TN 37235-1826, United States

stephen.lovett@wheaton.edu
Department of Mathematics and Computer Science,
Wheaton College, 501 College Avenue, Wheaton, IL 60187,
United States

mcmillan.matthew.i@gmail.com
St Catherine's College, Oxford OX1 3UJ, United Kingdom

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.787

A numerical investigation of level sets of extremal Sobolev functions

Stefan Juhnke and Jesse Ratzkin

(Communicated by Kenneth S. Berenhaut)

We investigate the level sets of extremal Sobolev functions. For $\Omega \subset \mathbb{R}^n$ and $1 \leq p < 2n/(n-2)$, these functions extremize the ratio $\|\nabla u\|_{L^2(\Omega)}/\|u\|_{L^p(\Omega)}$. We conjecture that as p increases, the extremal functions become more "peaked" (see the introduction below for a more precise statement), and present some numerical evidence to support this conjecture.

1. Introduction

Let $n \geq 2$ and let $\Omega \subset \mathbb{R}^n$ be a bounded domain with piecewise Lipschitz boundary, satisfying a uniform cone condition. One can associate a large variety of geometric and physical constants to Ω, such as volume, perimeter, diameter, inradius, the principal frequency $\lambda(\Omega)$, and torsional rigidity $P(\Omega)$ (which is also the maximal expected exit time of a standard Brownian particle). For more than a century, many mathematicians have investigated how all these quantities relate to each other; [Pólya and Szegő 1951] provides the best introduction to this topic, which remains very active today, with many open questions.

In the present paper we investigate the quantity

$$\mathcal{C}_p(\Omega) = \inf\left\{ \frac{\int_\Omega |\nabla u|^2 \, d\mu}{\left(\int_\Omega |u|^p \, d\mu\right)^{2/p}} : u \in W_0^{1,2}(\Omega), u \not\equiv 0 \right\}. \tag{1}$$

The constant $\mathcal{C}_p(\Omega)$ gives the best constant in the Sobolev embedding:

$$u \in W_0^{1,2}(\Omega) \;\Rightarrow\; \|u\|_{L^p(\Omega)} \leq \frac{1}{\sqrt{\mathcal{C}_p(\Omega)}} \|\nabla u\|_{L^2(\Omega)}.$$

By Rellich compactness, the infimum in (1) is finite, positive, and realized by an extremal function u_p^*, which we can take to be positive inside Ω (see, for instance,

MSC2010: primary 65N30; secondary 35J20.
Keywords: extremal Sobolev functions, semilinear elliptic PDE, distribution function.
Most of the work described below comes from Juhnke's honors dissertation, completed under the direction of Ratzkin. Ratzkin is partially supported by the National Research Foundation of South Africa.

[Gilbarg and Trudinger 2001; Sauvigny 2004; 2005]). The Euler–Lagrange equation for critical points of the ratio in (1) is

$$\Delta u + \Lambda u^{p-1} = 0, \quad u|_{\partial\Omega} = 0, \tag{2}$$

where Λ is the Lagrange multiplier. In the case that $u = u_p^*$ is an extremal function, a quick integration by parts argument shows that the Lagrange multiplier Λ is given by

$$\Lambda = \mathcal{C}_p(\Omega)\left(\int_\Omega (u_p^*)^p \, d\mu\right)^{(2-p)/p}.$$

It is worth remarking that in two cases the PDE (2) becomes linear: that of $p = 1$ and $p = 2$. In the case $p = 1$, we recover the torsional rigidity as $P(\Omega) = (\mathcal{C}_1(\Omega))^{-1}$, and in the case $p = 2$, we recover the principal frequency as $\lambda(\Omega) = \mathcal{C}_2(\Omega)$. These linear problems are both very well-studied, from a variety of perspectives, and the literature attached to each is huge. From this perspective, the second author and Tom Carroll began a research project several years ago, studying the variational problem (1) as it interpolates between torsional rigidity and principal frequency, and beyond. (See, for instance, [Carroll and Ratzkin 2011; 2012].) Primarily, we are interested in two central questions:

- Which of the properties of $P(\Omega)$ and $\lambda(\Omega)$ (and their extremal functions) also hold for $\mathcal{C}_p(\Omega)$ (and its extremal functions)?
- Can we track the behavior of $\mathcal{C}_p(\Omega)$ and its extremal function u_p^* as p varies?

Some of our investigations have led us conjecture the following.

Conjecture 1. *Let $n \geq 2$ and let $\Omega \subset \mathbb{R}^n$ be a bounded domain with piecewise Lipschitz boundary satisfying a uniform cone condition. Normalize the corresponding (positive) extremal function u_p^* so that*

$$\sup_{x\in\Omega}(u_p^*(x)) = 1,$$

and define the associated distribution function

$$\mu_p(t) = \big|\{x \in \Omega : u_p^*(x) > t\}\big|.$$

Then within the allowable range of exponents, we have

$$1 \leq p < q \;\Rightarrow\; \mu_p(t) > \mu_q(t) \quad \textit{for almost every } t \in (0, 1). \tag{3}$$

If $n = 2$, the allowable range of exponents is $1 \leq p < q$, and if $n \geq 3$, the allowable range of exponents is $1 \leq p < q < 2n/(n-2)$.

Below we will present some compelling numerical evidence in support of this conjecture. The remainder of the paper is structured as follows. In Section 2 we provide some context for our present investigation, and describe some of the related work present in the literature. In Section 3 we describe the numerical method we use, as well as its theoretical background, and we present our numerical results in Section 4. We conclude with a brief discussion of future work and unresolved questions in Section 5.

2. Related results

In this section we will highlight some related theorems about principal frequency, torsional rigidity, qualitative properties of extremal functions, and other quantities. The following is by no means an exhaustive list.

The distribution function μ_p is closely related to a variety of rearrangements of a generic test function u for (1). One can rearrange the function values of a positive function in a variety of ways, and different rearrangements will yield different results. One of the most well-used rearrangements is Schwarz symmetrization, where one replaces a positive function u on Ω with a radially symmetric, decreasing function u^* on B^*, a ball with the same volume as Ω. The rearrangement is defined to be equimeasurable with u:

$$|\{u > t\}| = |\{u^* > t\}| \quad \text{for almost every function value } t.$$

Krahn [1925] used Schwarz symmetrization to prove an inequality conjectured by Rayleigh in the late 1880s:

$$\lambda(\Omega) \geq \left(\frac{|\Omega|}{\omega_n}\right)^{-2/n} \lambda(\boldsymbol{B}), \tag{4}$$

where $\boldsymbol{B}$ is the unit ball in $\mathbb{R}^n$, and ω_n its volume. Moreover, equality can only occur in (4) if $\Omega = \boldsymbol{B}$ apart from a set of measure zero. In fact, it is straightforward to adapt Krahn's proof to show

$$|\Omega| = |\boldsymbol{B}| \ \Rightarrow\ \mathcal{C}_p(\Omega) \geq \mathcal{C}_p(\boldsymbol{B}), \tag{5}$$

with equality occurring if and only if $\Omega = \boldsymbol{B}$ apart from a set of measure zero (see [Carroll and Ratzkin 2011]). One can also use similar techniques to prove, for instance, that the square has the greatest torsional rigidity among all rhombi of the same area [Pólya 1948].

However, there is certainly a limit to the results one can prove using only Schwarz (or Steiner) symmetrization, and to go further one must apply new techniques. Among these, one can rearrange by weighted volume [Payne and Weinberger 1960; Ratzkin 2011; Hasnaoui and Hermi 2014], which works well for wedge-shaped domains. One can rearrange by powers of u, or (more generally) by some function of the level sets of u [Payne and Rayner 1972; 1973; Talenti 1976; Chiti 1982].

If one is combining domains using Minkowski addition, then the Minkowski supconvolution is a very useful tool [Colesanti et al. 2006].

All these techniques are successful, to varying degrees, when studying (1) for a *fixed* value of p. However, we are presently at a loss with regards to applying them when allowing p to vary. There are comparatively few results comparing the behavior of $\mathcal{C}_p(\Omega)$ and its extremals u_p^* for different values of p.

It is well known [Trudinger 1968] that as $p \to 2n/(n-2)$, the solutions u_p^* become arbitrarily peaked, and the distribution function $\mu_p(t)$ approaches 0 on the interval $(\epsilon, 1)$ for any $\epsilon > 0$. This behavior is a reflection of the fact that the Sobolev embedding is not compact for the critical exponent of $2n/(n-2)$, and the loss of compactness is due to the fact that the functional in (1) is invariant under conformal transformation for this exponent. Thus, it is interesting to understand the asymptotics as $p \to 2n/(n-2)$. A partial list of such results includes an asymptotic expansion of $\mathcal{C}_p(\Omega)$ due to van den Berg [2012] and a theorem of Flucher and Wei [1997] (see also [Bandle and Flucher 1996]) determining the asymptotic location of the maximum of the extremal u_p^*. Additionally, P. L. Lions [1984a; 1984b] started a program to understand the loss of compactness, due to concentration of solutions, for a variety of geometric problems in functional analysis and PDEs. R. Schoen and Y.-Y. Li (among others) have exploited this concentration-compactness phenomenon to understand the problem of prescribing the scalar curvature of a conformally flat metric.

We remark that until now we had scant evidence for Conjecture 1. Namely, we knew in advance that the extremals become arbitrarily peaked as p approaches the critical exponent, and we knew that in the very special case $\Omega = \boldsymbol{B}$, we have $\mu_1(t) > \mu_2(t)$.

3. Our numerical algorithm

Our numerical method is borrowed from foundational work of Choi and McKenna [1993] and Li and Zhou [2001], and its theoretical underpinning is the famous "mountain pass" method of Ambrosetti and Rabinowitz [1973]. Within our range of allowable exponents, Rellich compactness exactly implies that the functional (1) satisfies the Palais–Smale condition, and so the mountain pass theorem of [loc. cit.] implies the existence of a minimax critical point. A later refinement of Ni [1989] implies that in fact a minimax critical point lies on the Nehari manifold, defined by

$$\mathcal{M} = \left\{ u \in W_0^{1,2}(\Omega) : u \not\equiv 0, \int_\Omega |\nabla u|^2 - u^p \, d\mu = 0 \right\}. \tag{6}$$

To find critical points, we project onto $\mathcal{M}$, using the operator

$$P_{\mathcal{M}}(u) = \left(\frac{\int_\Omega |\nabla u|^2 \, d\mu}{\int_\Omega |u|^p \, d\mu} \right)^{1/(p-2)} u. \tag{7}$$

Our goal will be to find mountain pass critical points of the associated functional

$$\mathcal{I}(u) = \int_\Omega \tfrac{1}{2}|\nabla u|^2 - \tfrac{1}{p}|u|^p \, d\mu, \tag{8}$$

which lie on the Nehari manifold defined in (6). Observe that the Fréchet derivative of $\mathcal{I}$ is

$$\begin{aligned} \mathcal{I}'(u)(v) &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \mathcal{I}(u+\epsilon v) \\ &= \int_\Omega \langle \nabla u, \nabla v\rangle - u^{p-1} v \, d\mu, \end{aligned}$$

so that, after integrating by parts, we can find the direction v of steepest descent by solving the equation

$$2\lambda \Delta v = -\Delta u - u^{p-1}. \tag{9}$$

We are free to choose $\lambda > 0$ as a normalization constant, and choose it so that $\int_\Omega |\nabla v|^2 \, d\mu = 1$. (It is well known that by the Poincaré inequality this H^1-norm is equivalent to the $W^{1,2}$-norm.) An expansion of the difference quotient (using our normalization of v) shows

$$\frac{\mathcal{I}(u+\epsilon v) - \mathcal{I}(u)}{\epsilon} = -2\lambda + \mathcal{O}(\epsilon),$$

so choosing $\lambda > 0$ does indeed correspond to the direction of steepest descent of $\mathcal{I}$, rather than the direction of largest increase.

At this point we remark on the importance of taking $p > 2$. In the superlinear case, $u_0 \equiv 0$ is a local minimum and, so long as $u \not\equiv 0$, we have $\mathcal{I}(ku) < 0$ for $k > 0$ sufficiently large. Thus, for any path $\gamma(t)$ joining u_0 to ku_{guess}, the function $h_\gamma(t) = \mathcal{I}(\gamma(t))$ will have a maximum at some value t_γ. We can imagine varying the path γ and finding the lowest such maximal value, which is exactly our mountain pass critical point.

We will begin with an initial guess u_{guess} which is positive inside Ω and 0 on $\partial\Omega$, and let $u_1 = P_{\mathcal{M}}(u_{\text{guess}})$. Thereafter we apply the following algorithm:

(1) Given u_k, we compute the direction of steepest descent v_k using (9).

(2) If $\|v_k\|_{W^{1,2}(\Omega)}$ is sufficiently small, we stop the algorithm, and otherwise we let $u_{k+1} = P_{\mathcal{M}}(u_k + v_k)$

(3) If $\mathcal{I}(u_{k+1}) < \mathcal{I}(u_k)$ then we repeat the entire algorithm starting from the first step. Otherwise we replace v_k with $\frac{1}{2}v_k$ and recompute u_{k+1}.

(4) Upon the completion of this algorithm, we test our numerical solution to verify that it does indeed solve the PDE (2) weakly.

Several remarks are in order. The algorithm outlined above is exactly the one proposed by Li and Zhou [2001]. They proved convergence of the algorithm under a wide variety of hypotheses, which include the superlinear ($p > 2$) case of (1) and (8). However, they do not claim convergence of the algorithm in the sublinear case, and in this case the algorithm fails. On the other hand, we are able to verify that in the superlinear case the algorithm converges to a positive (weak) solution of the PDE (2), so we are confident we have reliable data in this case. We present this data in the next section.

In this algorithm we must repeatedly solve the linear PDE (9), which we do in the weak sense, using biquadratic (nine-noded) quadrilateral finite elements. In each of these steps we replace the corresponding integrals with sums over the corresponding elements. We outline this numerical step in the paragraphs below.

In this computation we take u as known at the mesh points (by an initial guess or by the result of a previous iteration). Writing $\bar{v} = 2\lambda v + u$, the solution to (9) is given by the solution to

$$\Delta \bar{v} = -u^{p-1}, \tag{10}$$

from which we can recover the steepest descent direction v.

To solve for $\bar{v} \in W_0^{1,2}(\Omega)$, we will solve the weak form of (10), i.e.,

$$\int_\Omega \nabla w(x) \cdot \nabla \bar{v}(x)\, dx = \int_\Omega w(x) u(x)^{p-1}\, dx \tag{11}$$

for any test function $w \in W_0^{1,2}(\Omega)$. We will now derive the finite element formulation based on the methods presented by Fish and Belytschko [2007]. We first notice that we can split up our integral as a sum of the integrals over the individual element domains Ω^e:

$$\sum_{e=1}^{n_{el}} \left(\int_{\Omega^e} \nabla w^e(x) \nabla \bar{v}^e(x)\, dx - \int_{\Omega^e} w^e(x) (\bar{v}^e(x))^{p-1}\, dx \right) = 0.$$

We now write our functions w and $\bar{v}$ in terms of their finite element approximations as

$$w(x) \approx w^h(x) = \boldsymbol{N}(x)\, \boldsymbol{w}, \qquad \bar{v}(x) \approx \bar{v}^h(x) = \boldsymbol{N}(x)\, \boldsymbol{d},$$

where $\boldsymbol{N}$ are quadratic shape functions with value 1 at their corresponding mesh point and value 0 at all other mesh points, while $\boldsymbol{w}$, $\boldsymbol{d}$ are vectors of nodal function values. The gradients of w and $\bar{v}$ can then be written as

$$\nabla w \approx \boldsymbol{B}(x)\, \boldsymbol{w}, \qquad \nabla \bar{v} \approx \boldsymbol{B}(x)\, \boldsymbol{d},$$

where $\boldsymbol{B}$ are the gradients of the shape functions. We can rewrite the above expressions for the element level as

$$w^e(x) \approx \boldsymbol{N}^e(x)\,\boldsymbol{w}^e, \quad \bar{v}^e(x) \approx \boldsymbol{N}^e(x)\boldsymbol{d}^e, \quad \nabla w^e \approx \boldsymbol{B}^e(x)\,\boldsymbol{w}^e, \quad \nabla \bar{v}^e \approx \boldsymbol{B}^e(x)\,\boldsymbol{d}^e.$$

Rewriting the integral using these approximations leaves us with

$$\sum_{e=1}^{n_{el}} \left(\int_{\Omega^e} \boldsymbol{w}^{e^T} \boldsymbol{B}^{e^T}(x)\boldsymbol{B}^e(x)\boldsymbol{d}^e \, dx - \int_{\Omega^e} \boldsymbol{w}^{e^T} \boldsymbol{N}^{e^T}(x)(\boldsymbol{N}^e(x)\boldsymbol{d}^e)^{p-1} \, dx \right) = 0,$$

since $(\boldsymbol{B}^e(x)\,\boldsymbol{w}^e)^T = \boldsymbol{w}^{e^T} \boldsymbol{B}^{e^T}(x)$ and $(\boldsymbol{N}^e(x)\,\boldsymbol{w}^e)^T = \boldsymbol{w}^{e^T} \boldsymbol{N}^{e^T}(x)$. We notice that we can take the constants $\boldsymbol{w}^{e^T}$ and $\boldsymbol{d}^e$ outside of the integral to give

$$\sum_{e=1}^{n_{el}} \boldsymbol{w}^{e^T} \left(\int_{\Omega^e} \boldsymbol{B}^{e^T}(x)\,\boldsymbol{B}^e(x)\,dx\,\boldsymbol{d}^e - \int_{\Omega^e} \boldsymbol{N}^{e^T}(x)(\boldsymbol{N}^e(x)\,\boldsymbol{d}^e)^{p-1} \, dx \right) = 0.$$

Letting

$$\boldsymbol{K}^e = \int_{\Omega^e} \boldsymbol{B}^{e^T}(x)\,\boldsymbol{B}^e(x)\,dx \quad \text{and} \quad \boldsymbol{f}^e = \int_{\Omega^e} \boldsymbol{N}^{e^T}(x)(\boldsymbol{N}^e(x)\,\boldsymbol{d}^e)^{p-1}\,dx$$

and using the gather matrix to write

$$\boldsymbol{w}^e = \boldsymbol{L}^e \boldsymbol{w}, \quad \boldsymbol{d}^e = \boldsymbol{L}^e \boldsymbol{d},$$

we get

$$\boldsymbol{w}^T \left(\sum_{e=1}^{n_{el}} \boldsymbol{L}^{e^T} \boldsymbol{K}^e \boldsymbol{L}^e \boldsymbol{d} - \sum_{e=1}^{n_{el}} \boldsymbol{L}^{e^T} \boldsymbol{f}^e \right) = 0.$$

Further letting

$$K = \sum_{e=1}^{n_{el}} \boldsymbol{L}^{e^T} \boldsymbol{K}^e \boldsymbol{L}^e \boldsymbol{d} \quad \text{and} \quad \boldsymbol{f} = \sum_{e=1}^{n_{el}} \boldsymbol{L}^{e^T} \boldsymbol{f}^e,$$

we end up with

$$\boldsymbol{w}^T(\boldsymbol{K}\boldsymbol{d} - \boldsymbol{f}) = 0, \quad \text{for all } \boldsymbol{w}.$$

Since we know that $w \in W_0^{1,2}$ is arbitrary, we therefore solve the discrete finite element form

$$\boldsymbol{K}\boldsymbol{d} = \boldsymbol{f}, \tag{12}$$

with $\boldsymbol{N}\boldsymbol{d}$ the finite element approximation to $\bar{v}$ from which we can recover the steepest descent direction v.

4. Numerical results

In this section we describe our numerical results. We implemented the algorithm described in Section 3 using Matlab, and all the figures displayed below come from this implementation.

We first implement our method on a unit ball of dimension four. In this case, the solution is radially symmetric, so we only need to solve an ODE. We display plots of these solutions and the corresponding distribution functions in Figure 1.

Observe that, as we expected, the distribution function appears to be monotone, and that as $p \to 4 = 2n/(n-2)$ the solution becomes arbitrarily concentrated at the origin.

We can verify that we are indeed finding solutions to the correct PDE. For the cases $p = 1$ and $p = 2$, we can compute the solutions analytically, and verify directly that our numerical solutions agree quite well. These are (up to a constant multiple)

$$u_1^*(r) = 1 - r^2, \quad u_2^*(r) = r^{(2-n)/2} J_{(n-2)/2}(j_{(n-2)/2} r),$$

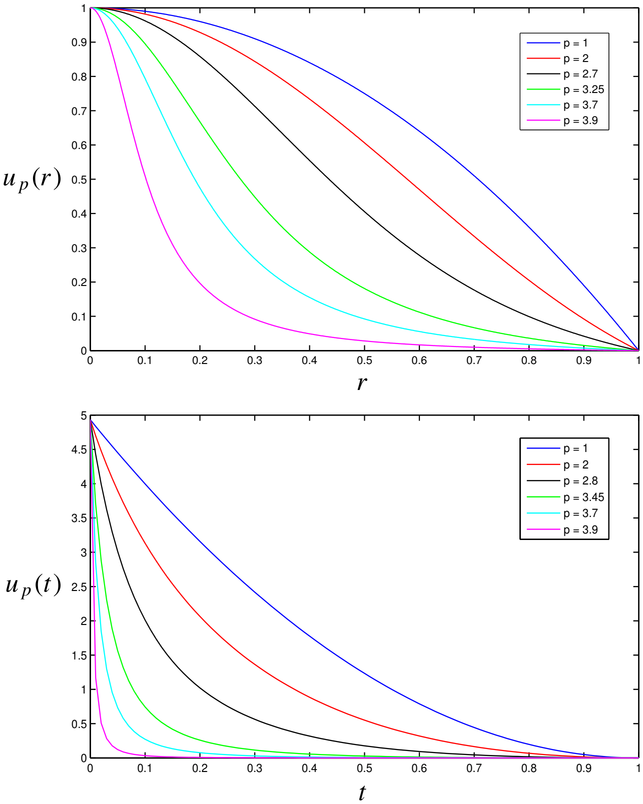


Figure 1. Extremal Sobolev functions (top) and their distributions (bottom) for a four-dimensional unit ball.

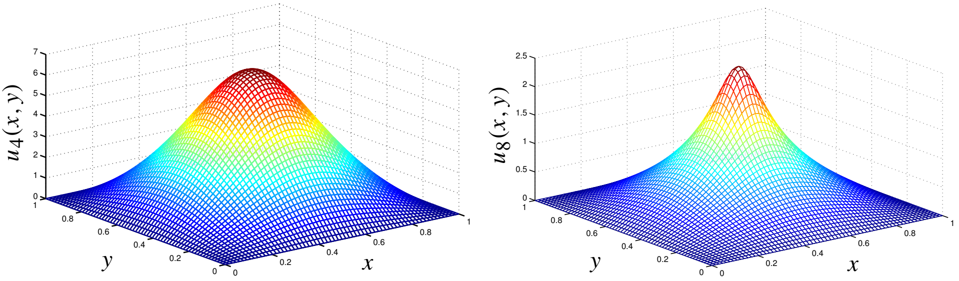


Figure 2. Extremal Sobolev functions for $p = 4$ (left) and $p = 8$ (right) on a unit square.

where J_a is the Bessel function of the first kind of index a and j_a is its first positive zero. For other values of p we can verify that we have found a weak solution of (2). As the solution is a priori radial, we know that the weak form of the PDE is

$$WT_w(u) := \int_0^1 \left(-r^{1-n} \frac{\partial w(r)}{\partial r} \left(r^{n-1} \frac{\partial u(r)}{\partial r} \right) + w(r) \Lambda u(r)^{p-1} \right) r^{n-1}\, dr = 0. \quad (13)$$

The above lends itself well to testing via finite element approximation. A random test function $w(r)$ is created by randomly generating numbers at the mesh points and $WT_w(u)$ is evaluated by Gauss quadrature. For comparison purposes, the functions u are normalized so that $\sup(u) = 1$. This requires that Λ be rescaled (Λ is set equal to 1 in the algorithm for simplicity), and the appropriate rescaling is then given by a^{2-p}, where a is the factor normalizing u. This rescaling is derived from the fact that if u solves

$$\Delta u + u^{p-1} = 0, \quad (14)$$

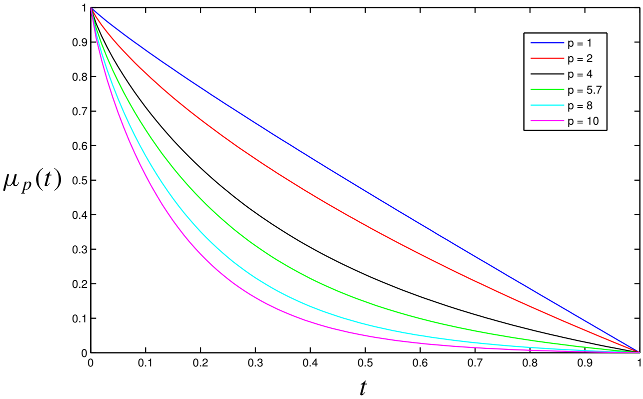


Figure 3. Distributions of extremal Sobolev functions for a unit square in the plane.

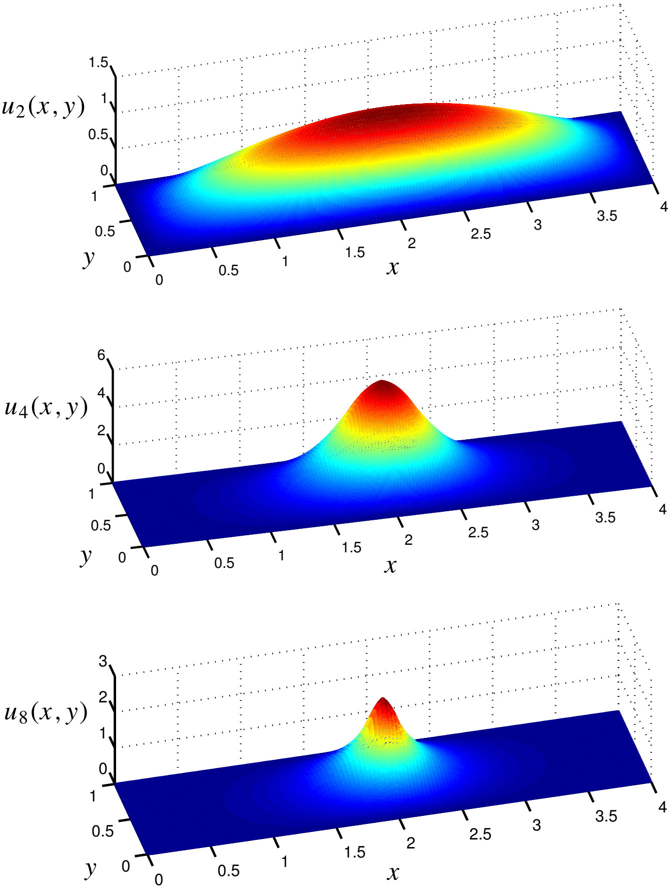


Figure 4. Extremal Sobolev functions for $p = 2$ (top), $p = 4$ (middle), and $p = 8$ (bottom) on a 1×4 rectangle.

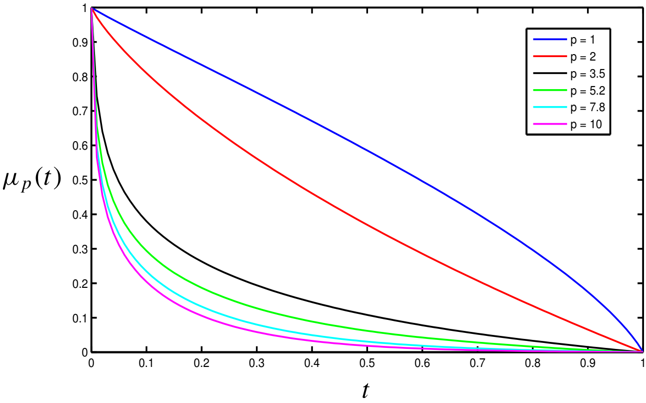


Figure 5. Distributions of extremal Sobolev functions for a 1×4 rectangle.

then au solves $\Delta(au) + a^{2-p}(au)^{p-1} = 0$, by simply multiplying (14) by a.

We generate values of $WT_w(u)$ for a number of test functions w and examine the average magnitude. As alluded to previously, the result of the test (13) is that for solution candidate functions derived from our algorithm for $2 \leq p < 2n/(n-2)$ and for $p = 1$, we have $WT_w(u)$ very close to zero, meaning that we can be confident that we have found appropriate solutions.

Next we implemented our algorithm in a unit square in the plane. We display plots of our numerical solutions for both $p = 4$ and $p = 8$ in Figure 2 and the distribution functions for several values of p in Figure 3. Again we verify that our numerical algorithm does find a weak solution of (2). This time we define

$$WT_w(u) := \int_\Omega \bigl(-\nabla u(x)\nabla w(x) + w\Lambda u(x)^{p-1}\bigr)\, dx \tag{15}$$

and again compute $WT_w(u)$ for our candidate solutions, with appropriate rescalings as described previously. We have closely matched the result of Choi and McKenna for the case $p = 4$, which means that we should be able to use the value $WT_w(u_4^*)$ as a gauge for how close to zero $WT_w(u)$ should be for appropriate solutions. Again we find that for $2 \leq p < 2n/(2-n)$ and $p = 1$, we get values of $WT_w(u)$ very close to zero and of the same magnitude as $WT_w(u_4^*)$.

Finally we implemented our algorithm on a rectangle of width 1 and length 4 in the plane. We display plots of our numerical solutions for $p = 2, 4$, and 8 in Figure 4, as well as the distribution functions for several values of p in Figure 5. We use the same test as we did in the case of the unit square to verify that in the case of the 1×4 rectangle, we have indeed found (weak) numerical solutions of (2).

5. Outlook

The present paper is only the start of our numerical and theoretical investigations into Conjecture 1. We would like to verify our results on some more planar domains, such as triangles and parallelograms. Next we anticipate numerical computations for higher dimensional objects, such as cubes and parallelepipeds, in the superlinear case, as well as possibly some ring domains. We will also need to develop a new numerical algorithm which yields reliable results for $1 < p < 2$. Finally, we hope that our numerical data provides enough insight to rigorously prove our conjecture.

References

[Ambrosetti and Rabinowitz 1973] A. Ambrosetti and P. H. Rabinowitz, "Dual variational methods in critical point theory and applications", *J. Funct. Anal.* **14** (1973), 349–381. MR 51 #6412 Zbl 0273.49063

[Bandle and Flucher 1996] C. Bandle and M. Flucher, "Harmonic radius and concentration of energy; hyperbolic radius and Liouville's equations $\Delta U = e^U$ and $\Delta U = U^{(n+2)/(n-2)}$", *SIAM Rev.* **38**:2 (1996), 191–238. MR 97b:35046 Zbl 0857.35034

[van den Berg 2012] M. van den Berg, "Estimates for the torsion function and Sobolev constants", *Potential Anal.* **36**:4 (2012), 607–616. MR 2904636 Zbl 1246.60108

[Carroll and Ratzkin 2011] T. Carroll and J. Ratzkin, "Interpolating between torsional rigidity and principal frequency", *J. Math. Anal. Appl.* **379**:2 (2011), 818–826. MR 2012d:35054 Zbl 1216.35016

[Carroll and Ratzkin 2012] T. Carroll and J. Ratzkin, "Two isoperimetric inequalities for the Sobolev constant", *Z. Angew. Math. Phys.* **63**:5 (2012), 855–863. MR 2991218 Zbl 1258.35154

[Chiti 1982] G. Chiti, "A reverse Hölder inequality for the eigenfunctions of linear second order elliptic operators", *Z. Angew. Math. Phys.* **33**:1 (1982), 143–148. MR 83i:35141 Zbl 0508.35063

[Choi and McKenna 1993] Y. S. Choi and P. J. McKenna, "A mountain pass method for the numerical solution of semilinear elliptic problems", *Nonlinear Anal.* **20**:4 (1993), 417–437. MR 94c:65133 Zbl 0779.35032

[Colesanti et al. 2006] A. Colesanti, P. Cuoghi, and P. Salani, "Brunn–Minkowski inequalities for two functionals involving the p-Laplace operator", *Appl. Anal.* **85**:1-3 (2006), 45–66. MR 2006j:52009 Zbl 1151.52307

[Fish and Belytschko 2007] J. Fish and T. Belytschko, *A first course in finite elements*, Wiley, Chichester, 2007. MR 2008d:74054 Zbl 1135.74001

[Flucher and Wei 1997] M. Flucher and J. Wei, "Semilinear Dirichlet problem with nearly critical exponent, asymptotic location of hot spots", *Manuscripta Math.* **94**:3 (1997), 337–346. MR 99b:35066 Zbl 0892.35061

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Springer, Berlin, 2001. Corrected 3rd printing. MR 2001k:35004 Zbl 1042.35002

[Hasnaoui and Hermi 2014] A. Hasnaoui and L. Hermi, "Isoperimetric inequalities for a wedge-like membrane", *Ann. Henri Poincaré* **15**:2 (2014), 369–406. MR 3159985 Zbl 06290599

[Krahn 1925] E. Krahn, "Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises", *Math. Ann.* **94**:1 (1925), 97–100. MR 1512244 JFM 51.0356.05

[Li and Zhou 2001] Y. Li and J. Zhou, "A minimax method for finding multiple critical points and its applications to semilinear PDEs", *SIAM J. Sci. Comput.* **23**:3 (2001), 840–865. MR 2002h:49012 Zbl 1002.35004

[Lions 1984a] P.-L. Lions, "The concentration-compactness principle in the calculus of variations. The locally compact case, part 1", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **1**:2 (1984), 109–145. MR 87e:49035a Zbl 0541.49009

[Lions 1984b] P.-L. Lions, "The concentration-compactness principle in the calculus of variations. The locally compact case, part 2", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **1**:4 (1984), 223–283. MR 87e:49035b Zbl 0704.49004

[Ni 1989] W.-M. Ni, "Recent progress in semilinear elliptic equations", *RIMS Kôkyûroku Bessatsu* **679** (1989), 1–39.

[Payne and Rayner 1972] L. E. Payne and M. E. Rayner, "An isoperimetric inequality for the first eigenfunction in the fixed membrane problem", *Z. Angew. Math. Phys.* **23** (1972), 13–15. MR 47 #2203 Zbl 0241.73080

[Payne and Rayner 1973] L. E. Payne and M. E. Rayner, "Some isoperimetric norm bounds for solutions of the Helmholtz equation", *Z. Angew. Math. Phys.* **24** (1973), 105–110. MR 48 #2554 Zbl 0256.35023

[Payne and Weinberger 1960] L. E. Payne and H. F. Weinberger, “A Faber–Krahn inequality for wedge-like membranes”, *J. Math. Phys.* (*MIT*) **39** (1960), 182–188. MR 23 #B1202 Zbl 0099.18701

[Pólya 1948] G. Pólya, “Torsional rigidity, principal frequency, electrostatic capacity and symmetrization”, *Quart. Appl. Math.* **6** (1948), 267–277. MR 10,206b Zbl 0037.25301

[Pólya and Szegő 1951] G. Pólya and G. Szegő, *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies **27**, Princeton University Press, 1951. MR 13,270d Zbl 0044.38301

[Ratzkin 2011] J. Ratzkin, “Eigenvalues of Euclidean wedge domains in higher dimensions”, *Calc. Var. Partial Differential Equations* **42**:1-2 (2011), 93–106. MR 2012m:35230 Zbl 1247.35080

[Sauvigny 2004] F. Sauvigny, *Partielle Differentialgleichungen der Geometrie und der Physik, 1: Grundlagen und Integraldarstellungen*, Springer, Berlin, 2004. Translated as *Partial differential equations, 1: Foundations and integral representations*, Springer, Berlin, 2006. 2nd ed published in 2012. MR 2007c:35001 Zbl 1049.35001

[Sauvigny 2005] F. Sauvigny, *Partielle Differentialgleichungen der Geometrie und der Physik, 2: Funktionalanalytische Lösungsmethoden*, Springer, Berlin, 2005. Translated as *Partial differential equations, 2: Functional analytic methods*, Springer, Berlin, 2006. 2nd ed published in 2012. MR 2007c:35002 Zbl 1072.35002

[Talenti 1976] G. Talenti, “Elliptic equations and rearrangements”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:4 (1976), 697–718. MR 58 #29170 Zbl 0341.35031

[Trudinger 1968] N. S. Trudinger, “Remarks concerning the conformal deformation of Riemannian structures on compact manifolds”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (3) **22** (1968), 265–274. MR 39 #2093 Zbl 0159.23801

Received: 2014-03-25 Revised: 2014-09-07 Accepted: 2014-10-01

juhnke.stefan@gmail.com

Department of Mathematics and Applied Mathematics,
University of Cape Town, Private Bag X1, Rondebosch,
Cape Town, 7701, South Africa

jesse.ratzkin@uct.ac.za

Department of Mathematics and Applied Mathematics,
University of Cape Town, Private Bag X1, Rondebosch,
Cape Town, 7701, South Africa

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.801

Coalitions and cliques in the school choice problem

Sinan Aksoy, Adam Azzam, Chaya Coppersmith,
Julie Glass, Gizem Karaali, Xueying Zhao and Xinjing Zhu

(Communicated by Kenneth S. Berenhaut)

The school choice problem (SCP) looks at assignment mechanisms matching students in a public school district to seats in district schools. The Gale–Shapley deferred acceptance mechanism applied to the SCP, known as the student optimal stable matching (SOSM), is the most efficient among stable mechanisms yielding a solution to the SCP. A more recent mechanism, the efficiency adjusted deferred acceptance mechanism (EADAM), aims to address the well-documented tension between efficiency and stability illustrated by SOSM. We introduce two alternative efficiency adjustments to SOSM, both of which necessarily sacrifice stability. Our discussion focuses on the mathematical novelty of new efficiency modifications rather than any practical superiority of implementation or outcome. That is, our contribution lies in process rather than outcome. Yet we argue that the demonstration of multiple processes yielding common outcomes is, in itself, a measure of the quality of that outcome. More specifically the consistency of outcome from different processes strengthens the argument that Pareto dominations of SOSM can be supported as "fair" despite the resulting priority violations.

1. Introduction

Since the mid-eighties, in cities across the United States, public school assignment policies have shifted towards providing students the opportunity to influence their school assignment. The main objective of these *school choice* policies is to allow all students to attend more desirable schools. A standard theoretical framework for studying such policies is two-sided matching (see [Gale 2001; Roth and Sotomayor 1990]). Presented in this context, the practical goal of the school choice problem (SCP) is to devise a matching mechanism (designed by or for the school district) that

MSC2010: 91B68, 90B80, 91B14.

Keywords: mechanism design, assignment, matching, school choice.

Aksoy, Azzam, Coppersmith, and Karaali were partially supported by NSF Grant DMS-0755540. Karaali was partially supported by NSA grant H98230-11-1-0186 and a Pomona College Hirsch Research Initiation Grant. Zhao was partially supported by the Hutchcroft Fund of the Department of Mathematics and Statistics at Mount Holyoke College. Zhu was partially supported by a Mount Holyoke College Ellen P. Reese Fellowship.

allocates available resources (seats in schools) among players (students or parents) subject to district priorities and legal requirements. Mathematically, it is interesting to consider ways in which the mechanisms might be modified that, while arguably consistent with the societal objectives of the SCP, present novel approaches to the underlying process.

In the economics literature, the SCP is viewed as a standard prototype for priority-based allocation problems (see [Kesten 2006]) and many of the school choice mechanisms in use or under investigation tolerate a large number of students receiving low preference schools (inefficiency) in order to respect school priority structures (stability). The ultimate purpose of these priorities is to benefit the students, but in many practical situations they are also the direct cause of efficiency losses, thus resulting in students receiving less desirable assignments than might have been possible. This suggests that taking a stable solution as a starting point and then making improvements for efficiency may be a reasonable compromise resulting in more desirable matchings.[1] Our discussion here will focus on the mathematical nuances of different efficiency modifications rather than any practical superiority of implementation or outcome. We will also suggest that stability loss may be justified to key stakeholders by arguing that the mathematical modifications are unbiased and incorporated as part of the overall process, and thus they do not constitute a “breach of contract”.

Throughout, we employ the language and methods of mechanism design as applied to the SCP following in the footsteps of, for example, [Abdulkadiroğlu and Sönmez 2003]. In this context the designer/principal is the school district (or whoever is choosing the mechanism to be used), students are the players, and schools are merely items to be consumed.

The Gale–Shapley deferred acceptance mechanism applied to the SCP, known as the student optimal stable matching (SOSM), is the most efficient among stable mechanisms yielding a solution to the SCP [loc. cit.]. In this article we examine two concrete processes that modify the outcome of SOSM and improve efficiency at the cost of stability. Our goal is to situate in a common framework a range of ideas introduced recently by several different authors, so that the mathematical connections between different outcomes and processes are more visible. More specifically, we focus here on using multiple cooperation/collaboration methods to obtain Pareto improvements of SOSM. We are interested in the process as well as the outcome and, in particular, we argue that examining multiple pathways strengthens the case for those outcomes both in theory and in practice.

[1] A relevant quote from [Abdulkadiroğlu et al. 2009]: “Pareto efficiency for the students is the primary welfare goal, but [. . .] stability of the matching, and strategyproofness in the elicitation of student preferences, are incentive constraints that likely have to be met for the system to produce substantial welfare gains over the [current] system.”

We begin in Section 1B by introducing two standard mechanisms used in this area of investigation: SOSM and its close neighbor, the efficiency adjusted deferred acceptance mechanism (EADAM), a more recently introduced mechanism which aims to address the well-documented tension between efficiency and stability illustrated by SOSM. In Section 2 we introduce the first of our approaches by studying the use of "coalitions" in order to modify SOSM school assignments. This section closely follows [Huang 2006], where it is shown that while the Gale–Shapley deferred acceptance algorithm (DA) disincentivizes strategic action by individuals, it is still feasible for groups to beat the system by coming together and strategizing. We adapt Huang's methods to the SCP and describe a process which we call the *coalition improvement procedure* in Section 2A. Using coalitions in the SCP allows us to approach efficiency modifications to SOSM in a new way and offers an alternative argument in support of previously known matching mechanisms. For example, this approach can result in the EADAM outcome along with other Pareto improvements of SOSM. We focus on properties of coalition improvements and comparisons to EADAM in Section 2B.

Following up on the coalition/cooperation theme, in Section 3 we introduce a second and related approach which focuses on groups of students who form trading cycles ("cliques") to improve their own assignments.[2] We examine the impact of these cliques as applied to the SOSM outcome. Once again, our approach deploys mathematical tools in a new context to produce several Pareto improvements on SOSM. We take the opportunity to show that the coalition improvements of Section 2 can also be integrated into this new framework, which proves to be a powerful construct to study cycle improvements of various kinds from a common point of view.

1A. ***Notation and basic terms used.*** Let I denote a nonempty set of students, and S a nonempty set of schools. A *matching* $M : I \to S \cup \{\text{null}\}$ is a function that associates every student $i \in I$ with exactly one school $M(i)$, or potentially no school at all, in which case $M(i) = \text{null}$. Write $\mathfrak{M}$ for the set of matchings. We will also occasionally want to talk about school quotas, which we will encode in a function $q : S \to \mathbb{N}$; in other words, for $s \in S$, $q(s)$ is the number of seats to be filled at school s.

A *preference profile* P_i for student $i \in I$ is a tuple $(S_1, \dots, S_n)$ where the S_j form a partition of S and every element of S_j is preferred to every element of S_k if and only if $j < k$.[3] Define the *ranking function* $\varphi_i : S \to \mathbb{N}$ of a student $i \in I$

[2] The term *clique* has a specific meaning in graph theory, unrelated to our work here.

[3] We will assume that student preference lists are complete, so it makes sense to define a preference list as a partition of the set of all schools. This is not always realistic however. Some students may wish to submit truncated lists, and this may or may not be allowed by school district policies. In fact, complete preference profiles in this context are rare. Often families are only permitted to list 3 to 7

by letting $\varphi_i(s)$ denote i's ranking of $s \in S$. In other words, $\varphi_i(s) = j$ if $s \in S_j$. If i prefers s_k to s_l, we write $s_k \succ_i s_l$, or simply $s_k \succ s_l$ if i is unambiguous. Note that the notation $\succ$ denotes a strict preference order; if we want to describe a weak order, we will write $\succeq$. We denote a set consisting of preference profiles for each student in I by $\boldsymbol{P} = \{P_i : i \in I\}$, and the space of all such sets is denoted by $\mathfrak{P}$.

A *priority structure* Π_s for school $s \in S$ is a tuple $(I_1, \dots, I_n)$ where the I_j form a partition of I and every element of I_j is preferred to every element of I_k if and only if $j < k$. If s prefers i_k to i_l, we write $i_k \succ_s i_l$, or simply $i_k \succ i_l$ if s is unambiguous. Once again, the notation $\succ$ denotes a strict preference order; if we want to describe a weak order, we will write $\succeq$. We denote a set consisting of priority structures for each school in S by $\boldsymbol{\Pi} = \{\Pi_s : s \in S\}$, and the space of all such complete sets is denoted by $\mathfrak{I}$.

A matching M' (*Pareto*) *dominates* M if $M'(i) \succeq_i M(i)$ for all i and $M'(j) \succ_j M(j)$ is strict for some j. A (*Pareto*) *efficient matching* is a matching that is not (Pareto) dominated.

A *matching mechanism* $\mathcal{M} : \mathfrak{P} \times \mathfrak{I} \rightarrow \mathfrak{M}$ is a function that takes an ordered pair $(\boldsymbol{P}, \boldsymbol{\Pi})$ of preferences and priorities and produces a matching.

Let Π_s be a priority structure for school s. A matching M *violates the priority* of $i \in I$ for s if there exist some $j \in I$ and $s' \in S$ such that

(1) $M(j) = s$, $M(i) = s'$: j gets assigned s under M and i gets assigned s' under M,

(2) $s \succ_i s'$: i prefers attending s over s', and

(3) $i \succ_s j$: s prioritizes i over j.

We say that a matching M is *stable* if

(1) M does not violate any priorities,

(2) no student is matched to a lower-ranked school when a more preferred school is unfilled, or more precisely, if $M(i) = s$, then for any school $s' \in S$ with $s' \succ_i s$, $\#\{j \in I \mid M(j) = s'\} = q(s')$,

(3) no student remains unmatched when a school is unfilled; that is, if $M(i) = \text{null}$, then for any school $s \in S$, $\#\{j \in I \mid M(j) = s\} = q(s)$.[4]

A *stable mechanism* is one that always produces stable matchings.

schools, choosing among many more. Then the district "completes" the student's profile, by first adding any school in her walk zone (if not already listed), and then "padding" the list with the schools that remain unlisted added to the end of the preference list, strictly below any listed by the student herself. Here we shall assume that when incomplete lists are allowed or unavoidable, the student preference lists are padded in this manner; our results will then work without modification.

[4] We assume that all students prefer being placed anywhere to being unassigned.

1B. ***Background.*** In this section we describe two well-studied mechanisms in the SCP: the student optimal stable matching (SOSM) and the efficiency adjusted deferred acceptance mechanism (EADAM). All mechanisms presented in this section use strict preference lists for students.

The first of these, SOSM, is based on the Gale–Shapley deferred acceptance algorithm (DA) [1962]. See [Roth and Sotomayor 1990] for an extensive review of the various applications of the DA algorithm and [Roth 2008] for a more recent historical overview. Gale and Shapley first described their method in the context of the *stable-marriage problem* (see [Knuth 1997]) and proposed applying it to the *college admissions problem*, a problem that in some ways resembles the SCP. Abdulkadiroğlu and Sönmez [2003] adapted the DA algorithm to the SCP and called it the student optimal stable mechanism (SOSM). Below is a brief description of this procedure.

Student optimal stable mechanism:

Round 1: Each student applies first to his or her first choice school. Each school then tentatively accepts the student(s) highest on its preference list among those who applied that round (such students are now waitlisted) and rejects the rest beyond its quota. We remove each waitlisted student from the market. All unwaitlisted students move on to the next round.

And in general:

Round k, $k \geq 1$: Each unassigned student applies to his or her next choice school. Each school considers the new applicants together with the current waitlist and repopulates the waitlist with those applicants who are highest on its priority list and rejects the rest beyond its quota. We remove each waitlisted student from the market. All unwaitlisted students move on to the next round. The algorithm runs until all students have been assigned.

SOSM performs well when evaluated for Pareto efficiency[5], stability, and strategyproofness and is viewed as a practical mechanism for implementation. In fact,

[5] We should qualify this assertion about the efficiency performance of SOSM. In the school choice problem as in many other matching markets, the preference and priority classes often are not singleton sets [Irving 1994; Manlove 2002]. In other words, there are many students in the same priority level for a given school, and it is conceivable that a student may wish to classify two or more schools in the same level of preference. The way SOSM and similar mechanisms deal with ties in such scenarios (often randomly and only on the school side, assuming students will submit strict preferences) creates arbitrary rankings, introduces artificial conditions, and results in a sizable efficiency loss (see [Erdil and Ergin 2008] for a study of tie-breaking in the school choice context and its efficiency cost). More generally it is known that many desirable properties of stable matching mechanisms are automatic only in the strict-preferences and strict-priorities scenario; once we allow indifferences, the problem often gets much more complicated [Manlove et al. 2002] and one might need to devise new goals and new extensions of the notion of stability (see [Chen 2012; Irving 1994]). We will say a bit more about indifferences in the final section of this paper.

several large districts such as New York City and Boston [Abdulkadiroğlu et al. 2009; 2006; 2005a; 2005b] have adopted SOSM as their mechanism of choice. As we have mentioned, SOSM offers a stable strategyproof mechanism whose outcomes Pareto dominate all other stable matchings.[6]

As motivation for investigating efficiency adjustments to the SOSM outcome and for introducing a powerful and well-respected model mechanism (EADAM), we give an example, due to Roth, that illustrates the problem of efficiency versus stability in SOSM (see [Abdulkadiroğlu et al. 2009; Kesten 2010]). This example also suggests that one could consider alternative processes that maintain appropriate respect for the (players') input while allowing for viable algorithmic alternatives.

Assume there are three schools, s_1, s_2, s_3 and three students i_1, i_2, i_3. The priorities of the schools and the preferences of the students are given by

$$\mathrm{SCP}_1:\quad \begin{array}{ll} i_1 : s_2 \succ s_1 \succ s_3, & s_1 : i_1 \succ i_3 \succ i_2, \\ i_2 : s_1 \succ s_2 \succ s_3, & s_2 : i_2 \succ i_1 \succ i_3, \\ i_3 : s_1 \succ s_2 \succ s_3, & s_3 : i_2 \succ i_1 \succ i_3, \end{array}$$

where $a \succ b$ stands for "*a is preferable to b*". Here, the only stable matching is

$$M_S^{\mathrm{SCP}_1} = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_1 & s_2 & s_3 \end{pmatrix},$$

but this matching is (Pareto) dominated by

$$M_E^{\mathrm{SCP}_1} = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}.$$

We see that $M_E^{\mathrm{SCP}_1}$ (Pareto) dominates $M_S^{\mathrm{SCP}_1}$ because it assigns i_1 and i_2 schools they prefer over their $M_S^{\mathrm{SCP}_1}$ assignment. Furthermore, $M_E^{\mathrm{SCP}_1}$ is (Pareto) efficient. However, the matching is no longer stable because i_2 is in the position of violating i_3's priority for s_1.

In part to address the weakness illustrated by the example above, Kesten [2010] proposed a new mechanism, and called it the efficiency adjusted deferred acceptance mechanism (EADAM). In order to understand EADAM, we must first define an *interrupter*. Let student i be one who is tentatively placed in a school s at some step t while running the SOSM, and rejected from it at some later step t'. If there

[6] Note that a stable mechanism can never really be strategyproof in the complete sense. More specifically, *no stable matching mechanism exists for which stating the true preferences is always a best response for every agent where all other agents state their true preferences* (see for instance [Roth and Sotomayor 1990, Corollary 4.5]). However the DA/SOSM is practically strategyproof as we only view the students as strategic players and the student optimality implies that there is no incentive for the students to misrepresent their preferences (see [Roth 1982]). This perspective does not take into account manipulation by schools in capacity (see [Sönmez 1997]) or preferences, see [Ehlers 2010] for recent work addressing these issues.

exists at least one other student who is rejected from school s after step $t-1$ and before step t', then we call student i an *interrupter* for school s and the pair (i,s) is an *interrupting pair* of step t'. An interrupter is *consenting* if she allows the mechanism to violate her priorities at no expense to her, that is, if she allows the mechanism to drop her from the running for schools she was an interruptor for, thus ignoring her priority standing with such schools. Note that the student's actual assignment would remain the same if not improve, and the consent would cost her nothing; by definition, she would not have been assigned to any school for which she was an interrupter in the first place.

EADAM then runs as follows:

Efficiency adjusted deferred acceptance mechanism:

Round 0: Run SOSM.

Round 1: Find the last step (of SOSM run in Round 0) at which a consenting interrupter is rejected from the school for which he/she is an interrupter. Identify all interrupting pairs in that step which contain a consenting interrupter. If there are no such pairs, then stop. Otherwise for each identified interrupting pair (i,s), remove school s from the preference list of student i without changing the relative order of the remaining schools. Rerun SOSM with the new preference profile for all such i until all students have been assigned.

And in general:

Round k, $k \geq 1$: Find the last step (of SOSM run in the previous round) at which a consenting interrupter is rejected from the school for which he/she is an interrupter. Identify all interrupting pairs in that step which contain a consenting interrupter. If there is no such pair, stop. Otherwise for each identified interrupting pair (i,s), remove school s from the preference list of student i without changing the relative order of the remaining schools. Rerun SOSM with the new preference profile until all students have been assigned.

In SCP$_1$, (i_3,s_1) is an interrupting pair and EADAM with the consent of i_3 outputs the Pareto efficient matching $M_E^{\mathrm{SCP}_1}$. Note that this result improves the assignments for i_1 and i_2 while leaving i_3 with the same assignment. This mechanism deploys a balanced approach to priorities and preferences and points towards the possibility of introducing alternative pathways to these outcomes, which leads us to our next section where we do just that.

2. Coalitions in the school choice problem

Huang [2006] discusses a weakness of the Gale–Shapley algorithm in the context of the stable marriage problem and introduces the idea of *coalition cheating in the marriage problem*. More specifically he shows that a coalition can be formed where some men, without forgoing their own Gale–Shapley stable matching assignment, can cheat (misrepresent their preferences) so that some other men marry women who are higher on their preference list.

In this section we apply these ideas to the school choice problem. In this way we develop an alternative process for improving on the SOSM outcome. In Section 2A we give some background and an example that will motivate Huang's construction. We then introduce the elements of what Huang calls cheating coalitions in the context of the SCP, and discuss some implementation issues. From here onward, we resist the use of the term "cheating" in this context because we believe that these coalitions could be systematically incorporated into the design of a mechanism since they improve outcomes for some with no adverse effects on others. If the goal is for a "benevolent" district mechanism, then improving efficiency beyond that of a stable matching (e.g., SOSM), might simply be a part of the process. In Section 2B we compare the possible outcomes of coalitions to that of EADAM. Our presentation and general approach here are consistent with our focus on process as the primary area of interest while maintaining loyalty to the practical needs of the SCP framework.

2A. ***Huang's construction, coalitions and school choice.*** The following theorem establishes that in the stable marriage problem, there exists no coalition of men that may falsify their preferences such that every member of the coalition receives a *strictly better* assignment:

Theorem 2.1 [Dubins and Freedman 1981]. *In the Gale–Shapley men-optimal algorithm, no subset of men can improve their assignment by falsifying their preference lists.*

Translating the stable marriage problem to the context of the SCP, as done in [Abdulkadiroğlu and Sönmez 2003] by replacing men with s students, we get as an immediate corollary:

Corollary 2.2. *In the SOSM algorithm, no subset of students can improve their assignment by falsifying their preference lists.*[7]

In light of results of this nature, Huang [2006] introduces a nuanced notion of coalitions that falsify preferences to improve assignments. In the following, we carry over to the SCP setting this coalition model, which distinguishes between two main groups of players: those who falsify their preferences, and those who benefit from the falsifications.

Let I and S be the set of students and schools respectively in a given SCP. Let M be the SOSM stable matching assignment for the case where all students submit their true preferences. A *coalition* C is defined in terms of a pair (K, A) of subsets of the

[7] One can nonetheless prove that SOSM (DA as applied to school choice) is not group-strategyproof. We choose not to go further into strategy discussions here.

set I of students. The first subset, the *cabal* $K = (i_1, i_2, \dots, i_{|K|})$ of a coalition C, is a list of students such that each student i_k, $1 \le k \le |K|$, prefers $M(i_{k+1})$ to $M(i_k)$, indices taken modulo $|K|$. In other words, we have $M(i_{k+1}) \succ_{i_k} M(i_k)$ for $1 \le k \le |K|$, and a *cabal loop*, written $(i_1 \to i_2 \to \cdots \to i_{|K|} \to i_1)$, a closed chain of students each of whom would prefer the stable assignment of the person following him to his own stable assignment. The second subset, the *accomplice set* $A = A(K)$ of cabal $K = (i_1, i_2, \dots, i_{|K|})$, is a set of students $A(K) \subset I$ such that $i \in A(K)$ if for some $i_k \in K$, we have $M(i_{k+1}) \succ_i M(i)$ and $i \succ_{M(i_{k+1})} i_k$. In other words, an *accomplice* is a student who in his truthful preference list ranks the stable assignment of someone in the cabal (i_{k+1}) higher than his own stable assignment, while he himself is ranked higher by that school than another member of the cabal (the one pointing toward i_{k+1}) who would prefer it to his own school. Note that K and $A(K)$ may or may not be disjoint.

For any student $i \in I$, we can write the preference profile of i as a disjoint union of three sets: $(P_L[i], M(i), P_R[i])$. Here the set $P_L[i]$ (respectively $P_R[i]$) is simply the list of schools on i's preference profile to the left (respectively to the right) of his stable assignment $M(i)$. Let π_r denote a random permutation of S. We can now prove the following (as an easy adaptation from the analogous result of Huang):

Theorem 2.3 (cf. [Huang 2006]). *Let M be the SOSM matching for a given SCP when students submit their true preferences. Consider a coalition $C = (K, A(K))$, and suppose that each accomplice $i \in A(K)$ submits a falsified list of the form $(\pi_r(P_L[i] - X), M(i), \pi_r(P_R[i] \cup X))$, where*

- *if $i \notin K$, then $X = \{s \in M(K) \mid s = M(i_k), s \succ_i M(i), i \succ_s i_{k-1}\}$, and*
- *if $i = i_k \in A(K) \cap K$, then*

$$X = \{s \in M(K) \mid s = M(i_j), j \neq k, s \succ_{i_k} M(i_k), i_k \succ_s i_{j-1}\}.$$

Then in the resulting matching M', $M'(i_k) = M(i_{k+1})$ for $i_k \in K$ and $M'(i) = M(i)$ for $i \notin K$.

We observe that accomplices modify their preference profiles by moving schools on the left of their stable assignment to the right of their stable assignment if they are desirable to other students in the cabal. In particular, if i is an accomplice, then the set X of schools i moves to the right of his stable assignment will consist of all the stable assignments of the members of the cabal that rank i higher than the student following their stable assignment in the cabal loop. Note that the falsified preference lists incorporate a random permutation π_r of the preferences to the left and the right of the stable partner. The coalition procedure is quite robust, in that such a random permutation will not affect the outcome. In other words, the resulting matching creates a cyclical reassignment of those within the cabal loop while leaving all other assignments as they were.

We call each outcome of the improvement process described in Theorem 2.3 a *coalition improvement* and formalize the concept in the coalition improvement procedure (CIP).

Coalition improvement procedure for a given sequence of coalitions $(C_1,\dots,C_k)$[8]:

Round 0: Given a preference and priority profile, run the SOSM algorithm and obtain a temporary matching M_0.

Round t, $1 \le t \le k$: Given M_{t-1}, apply Theorem 2.3 with the coalition $C_t = (K_t, A(K_t))$. Return the resulting matching M'_{t-1} as the outcome M_t.

Let us now consider an example, which we will label SCP$_2$. This example demonstrates what CIP might look like in practice and also points out which set of consenting interruptors would result in EADAM having the same outcome. Let $I = \{i_1, i_2, i_3, i_4, i_5\}$ and $S = \{s_1, s_2, s_3, s_4, s_5\}$ be the sets of students and schools, respectively, and let their respective preference and priority profiles be given as follows:

$$\text{SCP}_2: \quad \begin{array}{ll} i_1 : s_2 \succ s_5 \succ s_4 \succ s_3 \succ s_1, & s_1 : i_3 \succ i_2 \succ i_4 \succ i_1 \succ i_5, \\ i_2 : s_2 \succ s_5 \succ s_4 \succ s_1 \succ s_3, & s_2 : i_4 \succ i_5 \succ i_1 \succ i_2 \succ i_3, \\ i_3 : s_5 \succ s_2 \succ s_1 \succ s_3 \succ s_4, & s_3 : i_2 \succ i_3 \succ i_4 \succ i_5 \succ i_1, \\ i_4 : s_4 \succ s_1 \succ s_2 \succ s_3 \succ s_5, & s_4 : i_1 \succ i_2 \succ i_3 \succ i_5 \succ i_4, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ s_3 \succ s_1, & s_5 : i_1 \succ i_2 \succ i_5 \succ i_3 \succ i_4. \end{array}$$

Note that the matching output by SOSM for SCP$_2$ is

$$M_S^{\text{SCP}_2} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_5 & s_4 & s_1 & s_2 & s_3 \end{pmatrix}.$$

We now consider the following coalition $C = (K, A(K))$: Let $K = \{i_1, i_2, i_4\}$ with the cabal loop $(i_1 \to i_4 \to i_2 \to i_1)$. The accomplice set $A(K)$ is $\{i_5\}$ and the set X for i_5 is $\{s_2, s_4\}$. In other words, the only student who modifies his preference profile is i_5. We display his old and new profiles:

$$i_5\text{'s old profile}: s_5 \succ s_4 \succ s_2 \succ \underline{s_3} \succ s_1,$$
$$i_5\text{'s new profile}: s_5 \succ \underline{s_3} \succ s_1 \succ s_2 \succ s_4.$$

(We underlined i_5's stable assignment s_3.) The outcome matching when we rerun SOSM is

$$M_C^{\text{SCP}_2} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_2 & s_5 & s_1 & s_4 & s_3 \end{pmatrix},$$

[8] It should be apparent that there may be multiple outcomes of CIP for a given SCP depending on the particular sequence of coalitions we input. For simplicity we will assume that the cabals in each of the C_i are disjoint.

which improves the outcome for all members of the cabal and does not affect the remaining students. We note that this is also the EADAM outcome if i_5 consents. We will discuss this example further in Section 3A.

2B. ***Coalitions and EADAM.*** SOSM's strict adherence to stability and the resulting inefficiency has already been mentioned here (and documented in [Abdulkadiroğlu et al. 2009; Kesten 2010] and elsewhere). In this section we compare CIP (described in Section 2A) and EADAM from [Kesten 2010] (described in Section 1B), both of which model efficiency adjustments to SOSM. Specifically, we show that the common outcome of CIP and EADAM demonstrated by SCP$_2$ holds more generally by proving that for any SCP, there exists a coalition so that CIP yields the EADAM outcome with full consent. This fact may justify "fairness" arguments despite the sacrifice of stability.

Here is our general statement:

Theorem 2.4. *For any possible combination of consenters, the associated EADAM outcome may be obtained by forming an appropriately designed coalition and running CIP.*

The intuition behind this is that accomplices can be viewed as interrupters who consent to waive their priority so that they do not start a rejection chain. But coalitions reframe the argument so that the players are given the power to improve the outcome of the mechanism rather than being asked to waive their priority as with consenters.

Proof of Theorem 2.4. Let I and S be the sets of students and schools, respectively. Let $(\boldsymbol{P}, \boldsymbol{\Pi})$ be a given school choice problem for the pair (I, S), and let W be the set of students who consent to waiving their priorities under EADAM. Denote by M_S and M_E the SOSM and the EADAM outcome matchings of this problem, respectively. We will now construct a coalition C which will result in the same outcome M_E. First define the cabal set K to be the set of all students whose assignments are different under M_S and M_E:

$$K = \{i \in I \mid M_S(i) \neq M_E(i)\}.$$

These are the students who benefit from EADAM; they will also be the students who will benefit from the coalition C. Since every student whose assignment changes under EADAM is in K, we can partition K into cabal loops. This is equivalent to the basic algebraic fact that any finite permutation can be written as the product of disjoint cycles. Hence an elementary algorithm to decompose K into its individual cabal loops can be described as follows:

Step 0: Define a permutation π_K of K by setting $\pi_K(i') = i$ (i' *points to* i) if $M_S(i) = M_E(i')$. In words, *i' points to i if EADAM matches i' to the school to which SOSM matches i.*

Step 1: Pick a student $i \in K$ and label her $i_{1,1}$. Then let $i_{1,2}$ be the student $\pi_K(i_{1,1})$ and more generally label $i_{1,j+1} = \pi_K(i_{1,j})$. This process will stop at some j_1 with $\pi_K(i_{1,j_1}) = i_{1,1}$, as π_K is a finite permutation. Then

$$K_1 = (i_{1,1} \to i_{1,2} \to \cdots \to i_{1,j_1} \to i_{1,1})$$

is a cabal loop.

And in general:

Step k, $k \geq 1$: Pick a student $i \in K$ who has not yet been assigned to a cabal loop and label her $i_{k,1}$. If none exists then the algorithm stops. Otherwise, label $\pi_K(i_{k,1})$ as $i_{k,2}$ and more generally label $i_{k,j+1} = \pi_K(i_{k,j})$. This process stops at some j_k with $\pi_K(i_{k,j_k}) = i_{k,1}$ as π_K is finite. Then $K_k = (i_{k,1} \to i_{k,2} \to \cdots \to i_{k,j_k} \to i_{k,1})$ is a cabal loop.

Note that the algorithm has to stop because K is finite. Furthermore each student in K shows up in exactly one round and hence in exactly one cabal loop, because π_K is invertible.

Next we describe how to form the accomplice set $A(K)$. A student i will be in $A(K)$ if and only if the following two conditions are both satisfied:

- $i \in W$, or equivalently, i consents to waive her priorities in EADAM.
- There is a school s such that (i, s) is a last interrupter pair at some round of EADAM.

The new preference profile for an accomplice $i \in A(K)$ will be of the form

$$(P_L[i] - X, M_S(i), P_R[i] \cup X),$$

where

- if $i \notin K$, then $X = \{s \in M_S(K) \mid s = M_S(i_k), s \succ_i M_S(i), i \succ_s i_{k+1}\}$, and
- if $i = i_k \in A(K) \cap K$, then

$$X = \{s \in M_S(K) \mid s = M_S(i_j), j \neq k, s \succ_{i_k} M_S(i_k), i_k \succ_s i_{j+1}\}.$$

Here we are using the notation of Section 2A where $P_L[i]$ (respectively $P_R[i]$) is the list of schools on i's preference profile to the left (respectively to the right) of his stable assignment $M_S(i)$.

Finally Theorem 2.3 allows us to conclude that the outcome matching M_C of $C = (K, A(K))$ will be as follows: $M_C(i) = M_S(i)$ for all $i \notin K$, and $M_C(i_k) = M_S(i_{k+1})$ for i_k, i_{k+1} in some cabal loop K_j in K. But then $M_C = M_E$ and we are done. □

It is interesting to observe that CIP can produce outcomes that cannot be obtained via EADAM no matter which students consent. That is, the converse of Theorem 2.4 is not true. To see this we analyze a minor modification of SCP$_2$ which we label SCP$_3$. Let $I = \{i_1, i_2, i_3, i_4, i_5\}$ and $S = \{s_1, s_2, s_3, s_4, s_5\}$ be given with the following preference and priority structures, respectively:

$$\text{SCP}_3: \quad \begin{array}{ll} i_1 : s_1 \succ s_2 \succ s_5 \succ s_4 \succ s_3, & s_1 : i_3 \succ i_2 \succ i_4 \succ i_1 \succ i_5, \\ i_2 : s_2 \succ s_5 \succ s_4 \succ s_1 \succ s_3, & s_2 : i_4 \succ i_5 \succ i_1 \succ i_2 \succ i_3, \\ i_3 : s_5 \succ s_2 \succ s_1 \succ s_3 \succ s_4, & s_3 : i_2 \succ i_3 \succ i_4 \succ i_5 \succ i_1, \\ i_4 : s_4 \succ s_1 \succ s_2 \succ s_3 \succ s_5, & s_4 : i_1 \succ i_2 \succ i_3 \succ i_5 \succ i_4, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ s_3 \succ s_1, & s_5 : i_1 \succ i_2 \succ i_5 \succ i_3 \succ i_4. \end{array}$$

The SOSM outcome is

$$M_S^{\text{SCP}_3} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_5 & s_4 & s_1 & s_2 & s_3 \end{pmatrix}.$$

EADAM with full consent (in fact we only need i_5's consent) returns the matching

$$M_E^{\text{SCP}_3} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_1 & s_2 & s_5 & s_4 & s_3 \end{pmatrix}.$$

This corresponds to a coalition with the cabal set $\{i_1, i_2, i_3, i_4\}$ and the singleton accomplice set $\{i_5\}$. The set X for i_5 will be $X = \{s_2, s_4, s_5\}$. Note that there are two cabal loops: $(i_1 \to i_3 \to i_1)$ and $(i_2 \to i_4 \to i_2)$. There are indeed other coalitions that could be used for the same SCP. Take, for instance, the cabal to be $\{i_2, i_4\}$ and let $\{i_5\}$ be the singleton accomplice set. Then $X = \{s_2, s_4\}$ and we get

$$M_C^{\text{SCP}_3} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_5 & s_2 & s_1 & s_4 & s_3 \end{pmatrix}.$$

This outcome cannot be obtained via EADAM because once i_5 consents to waive his priorities, he has to consent fully, and all Pareto improvements involving the interrupter pairs he was a part of will also be made.

3. Cliques for school choice

EADAM and CIP provide us with ways to systematically improve upon SOSM matching. Both involve complicated procedures requiring the identification of problematic preference profiles (of interruptors or possible coalition members) and subsequent modification of preference profiles and/or priority violations. The ultimate goal in either case is the same: to Pareto improve upon SOSM in a way that justifies the resulting priority violation(s). In this section we propose another way to improve efficiency starting from the SOSM outcome. There are, again necessarily, priority violations in the final matching. The main idea is as follows: We begin by applying SOSM to the given SCP. Next, with no further consideration of priorities, we enter students into a trading market designed purely to improve school assignments from the point of view of student preferences.

In Section 3A, we describe in more detail our new theoretical approach, the *trading adjusted deferred acceptance procedure* (TADAP). While doing so, we explicitly associate a directed graph to a given matching to provide a visual tool to describe possible efficiency improvements. We investigate basic properties of TADAP and compare outcomes of TADAP with those of other methods in Section 3B. In particular, in keeping with our focus on process and the relationship between outcomes, we discuss how coalitions and cliques relate to one another and to other mechanisms involving cycle improvements. We also comment on implications for the school choice context.

3A. *The trading adjusted deferred acceptance procedure.* We now develop a systematic way to find all Pareto improvements upon a predetermined matching M in a given SCP. We will of course be particularly interested in the case where M is the outcome of SOSM.

We start by associating a directed weighted graph (V, E, w) to M as follows: Each student i is assigned a unique vertex v_i in V. There is an edge from vertex v_i to vertex v_j if student i desires student j's assignment under the given matching at least as much as, if not more than, the school to which he himself was assigned. An edge e from vertex v_i to vertex v_j has weight $w(e) = 0$ if student i desires student j's assignment under the given matching as much as, but not more than, the school to which he himself was assigned, and $w(e) = 1$ if the preference is strict.

In the above we can identify V with the set of students. With this in mind we now introduce the following:

Definition 3.1. Let I and S be a set of n students and a set of m schools, respectively, with respective preference and priority structures $(\boldsymbol{P}, \boldsymbol{\Pi})$. Let M be a matching for the associated SCP. We say that the directed weighted graph $G_M = (V, E, w)$ is the *(directed weighted) graph of the matching M* if $V = I$; for any pair of students (i, j), there is an edge e_{ij} from i to j if and only if $M(j) \succeq_i M(i)$; and for each edge $e_{ij} \in E$, $w(e_{ij}) = 0$ if $M(i) \succeq_i M(j)$, and $w(e_{ij}) = 1$ otherwise.

Using this terminology, we can make the following definition:

Definition 3.2 (cf. [Ergin 2002, Definition 1]). Let I, S, $(\boldsymbol{P}, \boldsymbol{\Pi})$, M and G_M be given as in Definition 3.1 and let $k \in \mathbb{N}$. A *clique of length k* consists of a sequence $(i_1, i_2, \dots, i_k)$ of k distinct students such that for each $s < k$, there is an edge in E from v_{i_s} to $v_{i_{s+1}}$, there is an edge in E connecting v_{i_k} back to v_{i_1}, and for some $s < k$, we have $w(e_{i_s,i_{s+1}}) = 1$ or $w(e_{i_k,i_1}) = 1$.[9] A similar cycle where $w = 0$ on

[9] What we call a *clique* is occasionally called a *trading cycle* in some of the literature. We use the former for brevity and also as a hint to the social context.

all edges is called a *null clique*. A matching whose graph contains no cliques (null or otherwise) is *acyclical*.

A straightforward result then follows:

Theorem 3.3. *If there exists a matching M (Pareto) dominating a matching M', then the directed graph $G_{M'}$ of M' admits a clique. Equivalently, if the directed graph of M' is acyclical, then M' is Pareto efficient. Conversely, if M' admits a clique, we can always find a matching M which Pareto dominates M' (equivalently, the directed graph of a Pareto efficient matching is acyclical).*[10]

Consider now the following procedure:

Trading adjusted deferred acceptance procedure:

Round 0: Given a preference and priority profile, run the SOSM algorithm and obtain a temporary matching M_0.

Round t, $t \geq 1$: Given M_{t-1}, consider the graph (V_t, E_t, w_t) of M_{t-1}. If there exists a student with no path through him, remove that student from the graph; his assignment under M_t will remain his assignment at the beginning of this round. If there are any cliques in the graph (V_t, E_t), pick one (note that different choices here may yield different results). For each edge from i to j in this clique, let M_t be the matching that assigns student i the school to which j was matched under M_{t-1}. If there is no clique, return M_{t-1} as the outcome M_t and stop.

It is apparent from the description above that there may be multiple outcomes of TADAP for a given SCP. In particular, in cases with multiple cliques, the procedure may output different matchings depending on which cycles are selected at rounds $t \geq 1$. Because following a clique yields a Pareto improvement, all outcomes of TADAP in which a nonempty clique exists will Pareto dominate the SOSM matching. In fact, any final outcome of TADAP will be Pareto efficient Pareto dominations of the initial SOSM matching. A district might choose to select cliques in an arbitrary manner and/or select cliques that include certain student populations over others (reinforcing their original priority structure) in order to define a trading adjusted deferred mechanism.

We begin with an example where the preference and priority structures are strict. (In such a situation, the weight function on the graph is uniformly 1 and can be ignored.) Consider once again SCP$_2$ (Section 2A) with five students and five schools

[10] In this theorem and in the rest of this section, we do not consider the case when there are some unassigned students and/or some unfilled places at a given school. If, on the other hand, this happens, some students can improve their assignment by taking a more preferred free place at a school without harming others. This means that a matching M may be Pareto dominated even in the case when the directed graph of M is acyclic; see [Abraham et al. 2005], where a necessary and sufficient condition for a matching to be Pareto optimal is proved.

each with one seat:

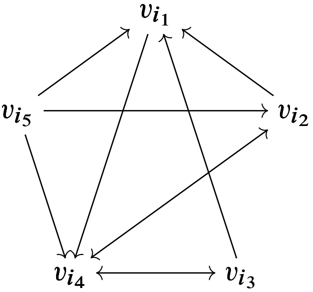
$$\text{SCP}_2: \quad \begin{array}{ll} i_1: s_2 \succ s_5 \succ s_4 \succ s_3 \succ s_1, & s_1: i_3 \succ i_2 \succ i_4 \succ i_1 \succ i_5, \\ i_2: s_2 \succ s_5 \succ s_4 \succ s_1 \succ s_3, & s_2: i_4 \succ i_5 \succ i_1 \succ i_2 \succ i_3, \\ i_3: s_5 \succ s_2 \succ s_1 \succ s_3 \succ s_4, & s_3: i_2 \succ i_3 \succ i_4 \succ i_5 \succ i_1, \\ i_4: s_4 \succ s_1 \succ s_2 \succ s_3 \succ s_5, & s_4: i_1 \succ i_2 \succ i_3 \succ i_5 \succ i_4, \\ i_5: s_5 \succ s_4 \succ s_2 \succ s_3 \succ s_1, & s_5: i_1 \succ i_2 \succ i_5 \succ i_3 \succ i_4. \end{array}$$

The matching under SOSM is

$$M_S^{\text{SCP}_2} = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_5 & s_4 & s_1 & s_2 & s_3 \end{pmatrix}.$$

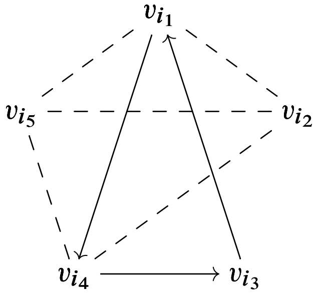
SOSM does a poor job with student preferences here. One student gets his fourth choice, three get their third choice and one gets his second choice.

For SCP$_2$, the associated SOSM matching can thus be translated into the following graph:

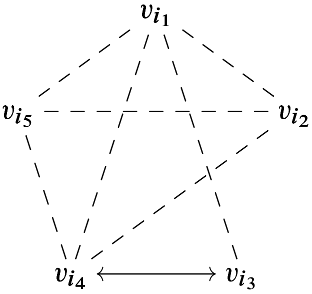


We see that if there is an arrow from i_l to i_j then i_l would (weakly) prefer to be assigned to $M(i_j)$. Such a swap can only be allowed if another student, i_k, prefers $M(i_l)$ to his own assignment, that is, only if there is a directed edge from some v_{i_k} to v_{i_l}. In this manner, a group of students can form a "swap market" and they can trade their SOSM assignments among themselves consistent with the directed graph. Such a swap market would correspond to a cycle in the graph. Here are four different cliques within the directed graph above (cliques denoted by unbroken arrows):

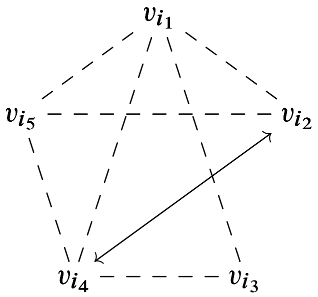
Cycle 1:



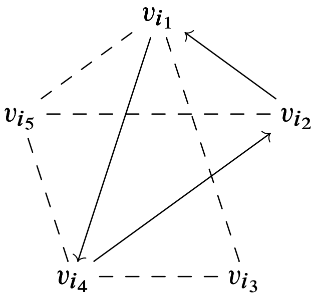
Cycle 2:



Cycle 3:



Cycle 4:



We list the assignments corresponding to each of the four cliques (note that the students' assignments are underlined in each matching):

$$M_1 = \begin{cases} i_1 : \underline{s_2} \succ s_5 \succ s_4 \succ s_3 \succ s_1, \\ i_2 : s_2 \succ s_5 \succ \underline{s_4} \succ s_1 \succ s_3, \\ i_3 : \underline{s_5} \succ s_2 \succ s_1 \succ s_3 \succ s_4, \\ i_4 : s_4 \succ \underline{s_1} \succ s_2 \succ s_3 \succ s_5, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ \underline{s_3} \succ s_1, \end{cases} \qquad M_2 = \begin{cases} i_1 : s_2 \succ \underline{s_5} \succ s_4 \succ s_3 \succ s_1, \\ i_2 : s_2 \succ s_5 \succ \underline{s_4} \succ s_1 \succ s_3, \\ i_3 : s_5 \succ \underline{s_2} \succ s_1 \succ s_3 \succ s_4, \\ i_4 : s_4 \succ \underline{s_1} \succ s_2 \succ s_3 \succ s_5, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ \underline{s_3} \succ s_1, \end{cases}$$

$$M_3 = \begin{cases} i_1 : s_2 \succ \underline{s_5} \succ s_4 \succ s_3 \succ s_1, \\ i_2 : \underline{s_2} \succ s_5 \succ s_4 \succ s_1 \succ s_3, \\ i_3 : s_5 \succ s_2 \succ \underline{s_1} \succ s_3 \succ s_4, \\ i_4 : \underline{s_4} \succ s_1 \succ s_2 \succ s_3 \succ s_5, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ \underline{s_3} \succ s_1, \end{cases} \qquad M_4 = \begin{cases} i_1 : \underline{s_2} \succ s_5 \succ s_4 \succ s_3 \succ s_1, \\ i_2 : s_2 \succ \underline{s_5} \succ s_4 \succ s_1 \succ s_3, \\ i_3 : s_5 \succ s_2 \succ \underline{s_1} \succ s_3 \succ s_4, \\ i_4 : \underline{s_4} \succ s_1 \succ s_2 \succ s_3 \succ s_5, \\ i_5 : s_5 \succ s_4 \succ s_2 \succ \underline{s_3} \succ s_1. \end{cases}$$

Observe that M_1, M_3, and M_4 are Pareto efficient but M_2 is not. In fact, if we draw the directed graph of M_2, we see that there is another cycle between i_3 and i_1. Thus we could continue with another clique, which would result in M_1. This raises the question of what efficient matching should be chosen in case of multiple efficient matchings. In this specific example, all three matchings give two students their top choice, one student her second choice, one student her third choice, and one student her fourth choice. Note that M_4 is the one obtained earlier via EADAM with the consent of i_5 and, equivalently, via a coalition with the cabal $K = \{i_1, i_2, i_4\}$ (the cabal loop is $(i_1 \to i_4 \to i_2 \to i_1)$), the accomplice set $A(K) = \{i_5\}$, and the set $X = \{s_2, s_4\}$ for i_5 (see Section 2A). One might argue that having multiple paths to a given outcome is, in itself, a justification to select that outcome as "best".

Note that in all these cases, i_5's assignment stays the same; in other words, i_5 can be labeled a "hopeless student" analogous to the "hopeless man" in [Huang 2006]. Looking at the graph, we see that there is no path passing through i_5; there is no chance for his situation to be improved. We can simplify the graph by taking out the vertex corresponding to i_5.

3B. ***Properties of TADAP.*** We begin this section with an analysis of the performance of TADAP under strategic action. We first state a key result from Kesten:

Proposition 3.4 [Kesten 2010, Proposition 4]. *No Pareto efficient mechanism that can Pareto improve upon SOSM is fully immune to strategic action.*

Since TADAP produces Pareto improvements of SOSM, it follows then that it is not strategyproof. This is consistent with other improvements upon SOSM. However, lack of strategyproofness does not imply easy manipulability. The feasibility of manipulation decreases as the size of the market (school district) increases. This is analogous to our earlier assertion that substantial coalitions are hard to form naturally on their own in the context of the SCP. Students do not have complete information about preference profiles of other students, so potential profitable strategic behaviors are highly unlikely. Formulating an alternative ranked list which yields a better assignment, even with complete information on all other students will most likely not be feasible for individual students.

Making the above more precise in technical language, we first split the schools into categories in terms of perceived quality. Then we can prove the following (cf. [Kesten 2010, Theorem 2]):

Theorem 3.5. *Let the set of schools S be partitioned into categories of perceived quality*

$$S = S_1 \cup S_2 \cup \cdots \cup S_m \quad \textit{with} \quad S_i \cap S_j = \varnothing \textit{ if } i \neq j$$

such that for any $k, l \in \{1, \ldots, m\}$ with $k < l$, each student prefers any school in S_k to any school in S_l. Let each student's information be symmetric for any two schools in the same perceived quality category. Then for any student, the strategy of truth telling stochastically dominates any other strategy when other students behave truthfully. Thus truth telling is an ordinal Bayesian Nash equilibrium of the preference revelation game under TADAP.

A well-studied method of strategic action by students is truncation manipulation, one of the few tools available in such a largely incomplete information matching game [Ehlers 2008]. However it is easy to see that in TADAP, no student benefits from truncating her preference list; any such truncation results in fewer cliques and fewer opportunities for that student (and for others) to improve her lot.

Note also that there is no strategy that a group of students could employ resulting in an outcome that is not among those produced by some choice of clique using TADAP. This is because in considering all possible cliques, we obtain all possible Pareto improvements.

Another prominent feature of TADAP is the efficiency of all its outcomes. Each clique followed improves the efficiency of the outcome, neutralizing to an extent the inefficiency caused by SOSM. As each such improvement creates a Pareto domination of the previous matching, at the end of the algorithm, we stop at a Pareto efficient matching. In fact, TADAP produces all efficient matchings that Pareto dominate SOSM. We can actually prove a slightly stronger result. A straightforward proof yields the following:

Proposition 3.6. *If matching M (Pareto) dominates the SOSM matching M^*, then M is realizable by TADAP up to null cliques.*

Obviously, distinct Pareto efficient matchings are Pareto incomparable. At this point we might resort to another evaluative criterion. For instance, we may wish to then consider the matchings with minimal preference index, a criterion that considers the sum of each player's priority violation as a measure of "lost utility";[11] this can reduce our option size. And, if the mechanism itself includes a second stage procedure such as TADAP or EADAM with full consent assumed, the overlap of outcomes may be called upon to justify the subsequent modification of outcomes. Since the overall process includes adjustments made in a standard manner to an initial stable outcome, the "fairness" is built in. If the standard adjustments are selected based upon criteria that include a "multiple pathway" argument, then the EADAM or other identified outcome is strongly supported. That is, no priority must be "waived" as that priority is part of the input, but needn't be incorporated into the final output matching.

The above proposition easily yields the following:

Corollary 3.7. *All efficient outcomes of EADAM and CIP can be found by TADAP.*

Recall that both EADAM and CIP provide us with efficiency improvements to SOSM. However, TADAP can return all Pareto efficient matchings that dominate SOSM so that we can compare all choices and pick the most desirable matching.

The absolute efficiency of TADAP may appeal to a utilitarian. However, this efficiency is achieved at the expense of stability. By its very construction, TADAP is not stable. Obviously we need to make an effort to coordinate the tradeoff between stability and efficiency. In the school choice literature, "fairness", "stability", "justified envy", and "no priority violation" are often used interchangeably. Here we propose a more nuanced notion of fairness (originally due to Kesten).

Since TADAP starts with the SOSM outcome as input, we are starting at a point where student priorities are considered and respected. TADAP may then make changes to the assignments which cause instability, manifesting itself in terms of justified envy. However, if a student's assigned school could not get any better under any stable mechanism, we surmise that his "justified envy" for anybody's assignment should not be justified. To formalize this we make the following definition:

[11] See [Aksoy et al. 2013; Karaali et al. 2012] for more on the preference index. Readers interested in other efficiency metrics might also refer to [Boudreau and Knoblauch 2010].

Definition 3.8 (cf. [Kesten 2010]). A matching is *reasonably fair* if there is no stable matching that can improve the assignment of any student. A mechanism is *reasonably fair* if it always outputs reasonably fair matchings.

Then the following is a direct consequence:

Proposition 3.9. *Matchings produced by TADAP are reasonably fair.*

Finally we should note that cycle improvements are used in the literature in a variety of ways. For instance Kesten [2010] describes such a model. In [Erdil and Ergin 2008], a stable cycle improvement model is developed. In this sense, the point of our work is to devise a scheme which incorporates any Pareto improvement of the SOSM outcome in a cycle improvement model.[12]

4. Conclusion

In this paper, we introduce and investigate the properties of coalitions and cliques, two notions that can be incorporated into a school choice mechanism to improve the efficiency of SOSM. Our focus is on the examination of mathematical processes for producing improvements. Both approaches we examine, coalitions and cliques, allow us to consider opportunities for cooperation and collaboration among and between the players and designers. We also hope that the mathematical tenor of our approach amidst a crowded literature focusing on practical outcomes will be aesthetically appealing and valuable for some readers.

The theoretical framework we are interested in might even have practical implications. We argue that the concerns about fairness that are prevalent in the literature of practically implementable mechanisms for school choice may be alleviated by our theoretical framework which demonstrates multiple pathways to produce outcomes of mechanisms commonly in use.

Our work may also be viewed as a fresh examination of two well-known and widely used school choice mechanisms (SOSM and EADAM). Our utilization of the notion of "reasonably fair" (originally proposed, to the best of our knowledge, by Kesten [2010]) captures our focus on cooperation and collaboration as a means to address any perceived unfairness. The double meaning of reasonableness as "somewhat" as well as "what a reasonable person would accept" is especially apropos. The constructions here yield opportunities to improve upon SOSM while justifying resulting priority violations in new ways.

Clearly our two modifications work by Pareto improving the baseline outcome of SOSM. Considering a coalition or clique improvement to SOSM as part of the

[12] Alternatively, rather than starting with a stable outcome and then modifying, one can start instead with an efficient outcome (such as one obtained via the top trading cycles mechanism) and then modify it to reach a more stable matching. Just such a method is investigated in [Morrill 2013].

overall mechanism with an established way of selecting the best overall outcome would allow for implementation without the need to establish approval from certain families. While it was not our goal here to develop a practical replacement for the well-established mechanisms now in use, we argue that the improvements presented here can have genuine practical implications. This is in part because of their coincidental outcomes rather than despite them. We can justify the priority violations that result from coalition improvement and cliques by showing that the new assignments (Pareto) dominate the SOSM assignments and can be arrived at via multiple paths. Because many of the current school priorities in place are meant to create some certainty/security for families, once those have been taken into account in the initial assignment, and since we can demonstrate that no families are made worse off, neither schools nor families should have a reason to object.

We also note that indifferences in student preferences may be incorporated into our model. Both collaborative approaches presented (coalitions and cliques) can work when students submit lists with indifferences. Although a considerable amount of research has been done regarding indifferences within school priority classes, indifference in student preferences has not been studied in as much depth. As far as we know, this characteristic of cycle improvement models has not been investigated before, at least in the school choice context. This can be a good avenue to pursue further.

As a final note, we once again emphasize the fact the two notions introduced in this paper are related to one another as well as to SOSM and EADAM. More specifically, given a coalition $C = (K, A(K))$ in the notation of Section 2A, we can always construct a sequence of cliques that under TADAP yields the same outcome. In other words, coalitional outcomes can always be obtained via TADAP as well. Going the other way is also doable in the case of strict preference profiles: any clique in such a context corresponds to a cabal cycle and the accomplices may be determined afterwards by looking at the resulting priority violations. It is precisely these overlapping and interlocking relationships between disparate processes that intrigues us and motivates this work.

References

[Abdulkadiroğlu and Sönmez 2003] A. Abdulkadiroğlu and T. Sönmez, "School choice: A mechanism design approach", *Am. Econ. Rev.* **93**:3 (2003), 729–747.

[Abdulkadiroğlu et al. 2005a] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth, "The New York City high school match", *Am. Econ. Rev.* **95**:2 (2005), 364–367.

[Abdulkadiroğlu et al. 2005b] A. Abdulkadiroğlu, P. A. Pathak, A. E. Roth, and T. Sönmez, "The Boston public school match", *Am. Econ. Rev.* **95**:2 (2005), 368–371.

[Abdulkadiroğlu et al. 2006] A. Abdulkadiroğlu, P. A. Pathak, A. E. Roth, and T. Sönmez, "Changing the Boston school-choice mechanism: Strategy-proofness as equal access", NBER Working Papers

11965, National Bureau of Economic Research, 2006, available at http://ideas.repec.org/p/nbr/nberwo/11965.html.

[Abdulkadiroğlu et al. 2009] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth, "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match", *Am. Econ. Rev.* **99**:5 (2009), 1954–1978.

[Abraham et al. 2005] D. J. Abraham, K. Cechlárová, D. F. Manlove, and K. Mehlhorn, "Pareto optimality in house allocation problems", pp. 1163–1175 in *Algorithms and computation*, edited by X. Deng and D.-Z. Du, Lecture Notes in Comput. Sci. **3827**, Springer, Berlin, 2005. MR 2258195 Zbl 1115.90049

[Aksoy et al. 2013] S. Aksoy, A. Azzam, C. Coppersmith, J. Glass, G. Karaali, X. Zhao, and X. Zhu, "School choice as a one-sided matching problem: Cardinal utilities and optimization", preprint, 2013. arXiv 1304.7413

[Boudreau and Knoblauch 2010] J. W. Boudreau and V. Knoblauch, "The price of stability in matching markets", Working papers 2010-16, University of Connecticut, Department of Economics, 2010, available at https://ideas.repec.org/p/uct/uconnp/2010-16.html.

[Chen 2012] N. Chen, "On computing Pareto stable assignments", pp. 384–395 in *29th International Symposium on Theoretical Aspects of Computer Science*, edited by C. Dürr and T. Wilke, LIPIcs. Leibniz Int. Proc. Inform. **14**, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2012. MR 2909330 Zbl 1245.91066

[Dubins and Freedman 1981] L. E. Dubins and D. A. Freedman, "Machiavelli and the Gale–Shapley algorithm", *Amer. Math. Monthly* **88**:7 (1981), 485–494. MR 82m:90089 Zbl 0449.92024

[Ehlers 2008] L. Ehlers, "Truncation strategies in matching markets", *Math. Oper. Res.* **33**:2 (2008), 327–335. MR 2009c:91107 Zbl 1231.91359

[Ehlers 2010] L. Ehlers, "Manipulation via capacities revisited", *Games Econom. Behav.* **69**:2 (2010), 302–311. MR 2011j:91215 Zbl 1230.91144

[Erdil and Ergin 2008] A. Erdil and H. I. Ergin, "What's the matter with tie-breaking? Improving efficiency in school choice", *Am. Econ. Rev.* **98**:3 (2008), 669–689.

[Ergin 2002] H. I. Ergin, "Efficient resource allocation on the basis of priorities", *Econometrica* **70**:6 (2002), 2489–2497. MR 2003k:91086 Zbl 1141.91563

[Gale 2001] D. Gale, "The two-sided matching problem: Origin, development and current issues", *Int. Game Theory Rev.* **3**:2-3 (2001), 237–252. MR 2002g:91004 Zbl 1127.91372

[Gale and Shapley 1962] D. Gale and L. S. Shapley, "College admissions and the stability of marriage", *Amer. Math. Monthly* **69**:1 (1962), 9–15. MR 1531503 Zbl 0109.24403

[Huang 2006] C.-C. Huang, "Cheating by men in the Gale–Shapley stable matching algorithm", pp. 418–431 in *Algorithms—ESA 2006*, edited by Y. Azar and T. Erlebach, Lecture Notes in Comput. Sci. **4168**, Springer, Berlin, 2006. MR 2347162 Zbl 1131.05319

[Irving 1994] R. W. Irving, "Stable marriage and indifference", *Discrete Appl. Math.* **48**:3 (1994), 261–272. MR 95b:90166 Zbl 0796.05078

[Karaali et al. 2012] G. Karaali, S. Aksoy, A. Azzam, C. Coppersmith, J. Glass, X. Zhao, and X. Zhu, "A cost-minimizing algorithm for school choice", in *Proceedings of the Twelfth International Symposium on Artificial Intelligence and Mathematics* (Fort Lauderdale, FL, 2012), edited by R. H. Sloan, 2012.

[Kesten 2006] O. Kesten, "On two competing mechanisms for priority-based allocation problems", *J. Econ. Theory* **127**:1 (2006), 155–171. MR 2006k:91036 Zbl 1125.91019

[Kesten 2010] O. Kesten, "School choice with consent", *Q. J. Econ.* **125**:3 (2010), 1297–1348. Zbl 1197.91153

[Knuth 1997] D. E. Knuth, *Stable marriage and its relation to other combinatorial problems: An introduction to the mathematical analysis of algorithms*, CRM Proceedings & Lecture Notes **10**, Amer. Math. Soc., Providence, RI, 1997. MR 97h:68048 Zbl 0860.68054

[Manlove 2002] D. F. Manlove, "The structure of stable marriage with indifference", *Discrete Appl. Math.* **122**:1-3 (2002), 167–181. MR 2003c:05004 Zbl 1008.05117

[Manlove et al. 2002] D. F. Manlove, R. W. Irving, K. Iwama, S. Miyazaki, and Y. Morita, "Hard variants of stable marriage", *Theoret. Comput. Sci.* **276**:1-2 (2002), 261–279. MR 2003b:05148 Zbl 1050.68171

[Morrill 2013] T. Morrill, "Making efficient school assignment fairer", preprint, 2013, available at http://www4.ncsu.edu/~tsmorril/papers/PrioritizedTradingCycles.pdf.

[Roth 1982] A. E. Roth, "The economics of matching: Stability and incentives", *Math. Oper. Res.* **7**:4 (1982), 617–628. MR 84f:90008

[Roth 2008] A. E. Roth, "Deferred acceptance algorithms: History, theory, practice, and open questions", *Internat. J. Game Theory* **36**:3-4 (2008), 537–569. MR 2009a:91081 Zbl 1142.91049

[Roth and Sotomayor 1990] A. E. Roth and M. A. O. Sotomayor, *Two-sided matching: A study in game-theoretic modeling and analysis*, Econometric Society Monographs **18**, Cambridge University Press, 1990. MR 93b:90001 Zbl 0726.90003

[Sönmez 1997] T. Sönmez, "Manipulation via capacities in two-sided matching markets", *J. Econom. Theory* **77**:1 (1997), 197–204. MR 98g:90055 Zbl 0892.90011

Received: 2014-04-09 Revised: 2014-09-22 Accepted: 2014-10-15

saksoy@ucsd.edu
Department of Mathematics,
University of California, San Diego,
9500 Gilman Drive # 0112, La Jolla, CA 92093, United States

adamazzam@math.ucla.edu
University of California, Los Angeles, Los Angeles, CA 95155,
United States

ccoppersmi@brynmawr.edu
Bryn Mawr College, Bryn Mawr, PA 19010, United States

julie.glass@csueastbay.edu
Department of Mathematics and Computer Science,
California State University, East Bay, RO 232,
Hayward, CA 94542, United States

gizem.karaali@pomona.edu
Department of Mathematics, Pomona College, 610 North
College Avenue, Claremont, CA 91711, United States

xueyingzhao2018@u.northwestern.edu
Northwestern University, Evanston, IL 60201, United States

xinjingzhu@gmail.com
Mount Holyoke College, South Hadley, MA 01075,
United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.825

The chromatic polynomials of signed Petersen graphs

Matthias Beck, Erika Meza, Bryan Nevarez, Alana Shine and Michael Young

(Communicated by Kenneth S. Berenhaut)

Zaslavsky proved in 2012 that, up to switching isomorphism, there are six different signed Petersen graphs and that they can be told apart by their chromatic polynomials, by showing that the latter give distinct results when evaluated at 3. He conjectured that the six different signed Petersen graphs also have distinct zero-free chromatic polynomials, and that both types of chromatic polynomials have distinct evaluations at *any* positive integer. We developed and executed a computer program (running in SAGE) that efficiently determines the number of proper k-colorings for a given signed graph; our computations for the signed Petersen graphs confirm Zaslavsky's conjecture. We also computed the chromatic polynomials of all signed complete graphs with up to five vertices.

Graph coloring problems are ubiquitous in many areas within and outside of mathematics. We are interested in certain enumerative questions about coloring signed graphs. A *signed graph* $\Sigma = (\Gamma, \sigma)$ consists of a graph $\Gamma = (V, E)$ and a signature $\sigma \in \{\pm\}^E$. The underlying graph Γ may have multiple edges and, besides the usual links and loops, also *half-edges* (with only one endpoint) and *loose edges* (no endpoints); the last are irrelevant for coloring questions, and so we assume in this paper that Σ has no loose edges. An unsigned graph can be realized by a signed graph all of whose edges are labeled with $+$. Signed graphs originated in the social sciences and have found applications also in biology, physics, computer science, and economics; see [Zaslavsky 1998–2012] for a comprehensive bibliography.

MSC2010: primary 05C22; secondary 05A15, 05C15.

Keywords: signed graph, Petersen graph, complete graph, chromatic polynomial, zero-free chromatic polynomial.

We are grateful to Thomas Zaslavsky and an anonymous referee for comments on an earlier version of this paper, and we thank Ricardo Cortez and the staff at MSRI for creating an ideal research environment at MSRI-UP. This research was partially supported by the NSF through the grants DMS-1162638 (Beck), DMS-0946431 (Young), and DMS-1156499 (MSRI-UP REU), and by the NSA through grant H98230-11-1-0213.

The *chromatic polynomial* $c_\Sigma(2k+1)$ counts the *proper k-colorings*

$$\boldsymbol{x} \in \{0, \pm 1, \dots, \pm k\}^V,$$

namely, those colorings that satisfy

$$x_v \neq \sigma_{vw}\, x_w$$

for any edge $vw \in E$ and $x_v \neq 0$ for any $v \in V$ incident with some half-edge. Zaslavsky [1982a] proved that $c_\Sigma(2k+1)$ is indeed a polynomial in k. It comes with a companion, the *zero-free chromatic polynomial* $c^*_\Sigma(2k)$, which counts all proper k-colorings $\boldsymbol{x} \in \{\pm 1, \dots, \pm k\}^V$.

The *Petersen graph* has served as a reference point for many proposed results in graph theory. Considering *signed* Petersen graphs, Zaslavsky [2012] showed that, while there are 2^{15} ways to assign a signature to the fifteen edges, only six of these are different up to switching isomorphism (a notion that we will make precise below), depicted in Figure 1. (In our figures we represent a positive edge with a solid line and a negative edge with a dashed line.)

Zaslavsky [2012] proved that these six signed Petersen graphs have distinct chromatic polynomials; thus they can be distinguished by this signed-graph invariant. He did not compute the chromatic polynomials but showed that they evaluate to distinct numbers at 3 [loc. cit., Table 9.2]. He conjectured that the six different signed Petersen graphs also have distinct zero-free chromatic polynomials, and that both types of chromatic polynomials have distinct evaluations at *any* positive integer [loc. cit., Conjecture 9.1]. Our first result confirms this conjecture.

Theorem 1. *The chromatic polynomials of the signed Petersen graphs* (denoted by $P_1, \dots, P_6$ in Figure 1) *are*

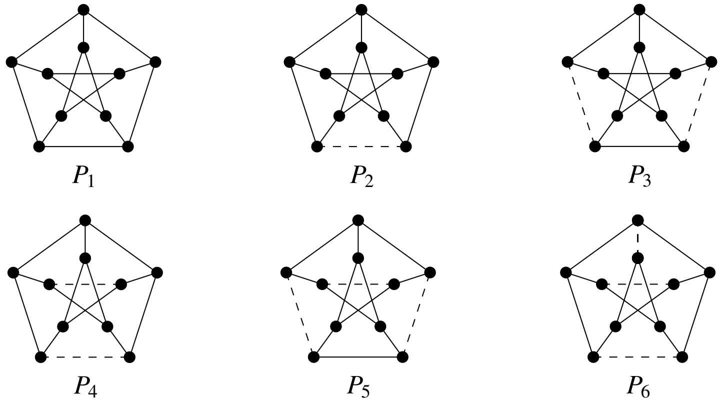


Figure 1. The six switching-distinct signed Petersen graphs.

$$\begin{aligned}
c_{P_1}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 3712k^6 \\
&\qquad - 1792k^5 + 160k^4 + 480k^3 - 336k^2 + 72k, \\
c_{P_2}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 3968k^6 \\
&\qquad - 2560k^5 + 1184k^4 - 352k^3 + 48k^2, \\
c_{P_3}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 4096k^6 \\
&\qquad - 2944k^5 + 1696k^4 - 760k^3 + 236k^2 - 40k, \\
c_{P_4}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 4224k^6 \\
&\qquad - 3200k^5 + 1984k^4 - 952k^3 + 308k^2 - 52k, \\
c_{P_5}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 4096k^6 \\
&\qquad - 3072k^5 + 1920k^4 - 960k^3 + 320k^2 - 48k, \\
c_{P_6}(2k+1) &= 1024k^{10} - 2560k^9 + 3840k^8 - 4480k^7 + 4480k^6 \\
&\qquad - 3712k^5 + 2560k^4 - 1320k^3 + 460k^2 - 90k.
\end{aligned}$$

Their zero-free counterparts are

$$\begin{aligned}
c^*_{P_1}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 86592k^6 \\
&\qquad - 91552k^5 + 68400k^4 - 34440k^3 + 10424k^2 - 1408k, \\
c^*_{P_2}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 86848k^6 \\
&\qquad - 93088k^5 + 72304k^4 - 39880k^3 + 14792k^2 - 3288k, \\
c^*_{P_3}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 86976k^6 \\
&\qquad - 93856k^5 + 74256k^4 - 42592k^3 + 16960k^2 - 4222k, \\
c^*_{P_4}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 87104k^6 \\
&\qquad - 94496k^5 + 75664k^4 - 44320k^3 + 18192k^2 - 4698k, \\
c^*_{P_5}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 86976k^6 \\
&\qquad - 93984k^5 + 74800k^4 - 43560k^3 + 17840k^2 - 4616k, \\
c^*_{P_6}(2k) &= 1024k^{10} - 7680k^9 + 26880k^8 - 58240k^7 + 87360k^6 \\
&\qquad - 95776k^5 + 78480k^4 - 47760k^3 + 20640k^2 - 5660k.
\end{aligned}$$

Consequently (*as a quick computation with a computer algebra system shows*), *none of the difference polynomials* $c_{P_m}(2k+1) - c_{P_n}(2k+1)$ *and* $c^*_{P_m}(2k) - c^*_{P_n}(2k)$, *with* $m \neq n$, *have a positive integer root.*

To compute the above polynomials, we developed and executed a computer program (running in SAGE [Stein et al. 2012]) that efficiently determines the number of proper k-colorings for any signed graph. This code can be downloaded from math.sfsu.edu/beck/papers/signedpetersen.sage or from the online supplement to this paper. The procedure `chrom` is the main method; it takes an incidence matrix and outputs the chromatic polynomial as an expression.

We also used our program to compute the chromatic polynomials of all signed complete graphs up to five vertices; up to switching isomorphism, there are two signed K_3s, three signed K_4s, and seven signed K_5s. As with the signed Petersen graphs, the chromatic polynomials distinguish these signed complete graphs:

Theorem 2. *The chromatic polynomials of the signed complete graphs* (denoted $K_3^{(1)}, K_3^{(2)}, \dots, K_5^{(7)}$ in Figure 2) *are*

$$\begin{aligned}
c_{K_3^{(1)}}(2k+1) &= 8k^3-2k,\\
c_{K_3^{(2)}}(2k+1) &= 8k^3,\\
c_{K_4^{(1)}}(2k+1) &= 16k^4-16k^3-4k^2+4k,\\
c_{K_4^{(2)}}(2k+1) &= 16k^4-16k^3+4k^2,\\
c_{K_4^{(3)}}(2k+1) &= 16k^4-16k^3+12k^2-2k,\\
c_{K_5^{(1)}}(2k+1) &= 32k^5-80k^4+40k^3+20k^2-12k,\\
c_{K_5^{(2)}}(2k+1) &= 32k^5-80k^4+64k^3-16k^2,\\
c_{K_5^{(3)}}(2k+1) &= 32k^5-80k^4+88k^3-48k^2+10k,\\
c_{K_5^{(4)}}(2k+1) &= 32k^5-80k^4+72k^3-28k^2+4k.\\
c_{K_5^{(5)}}(2k+1) &= 32k^5-80k^4+96k^3-56k^2+12k,\\
c_{K_5^{(6)}}(2k+1) &= 32k^5-80k^4+80k^3-40k^2+8k,\\
c_{K_5^{(7)}}(2k+1) &= 32k^5-80k^4+120k^3-80k^2+20k.
\end{aligned}$$

The corresponding zero-free chromatic polynomials are

$$\begin{aligned}
c^*_{K_3^{(1)}}(2k) &= 8k^3-12k^2+4k,\\
c^*_{K_3^{(2)}}(2k) &= 8k^3-12k^2+6k,\\
c^*_{K_4^{(1)}}(2k) &= 16k^4-48k^3+44k^2-12k,\\
c^*_{K_4^{(2)}}(2k) &= 16k^4-48k^3+52k^2-24k,\\
c^*_{K_4^{(3)}}(2k) &= 16k^4-48k^3+60k^2-34k,\\
c^*_{K_5^{(1)}}(2k) &= 32k^5-160k^4+280k^3-200k^2+48k,\\
c^*_{K_5^{(2)}}(2k) &= 32k^5-160k^4+304k^3-272k^2+114k,\\
c^*_{K_5^{(3)}}(2k) &= 32k^5-160k^4+328k^3-340k^2+174k,\\
c^*_{K_5^{(4)}}(2k) &= 32k^5-160k^4+312k^3-296k^2+136k,\\
c^*_{K_5^{(5)}}(2k) &= 32k^5-160k^4+336k^3-360k^2+190k,\\
c^*_{K_5^{(6)}}(2k) &= 32k^5-160k^4+320k^3-320k^2+158k,\\
c^*_{K_5^{(7)}}(2k) &= 32k^5-160k^4+360k^3-420k^2+240k.
\end{aligned}$$

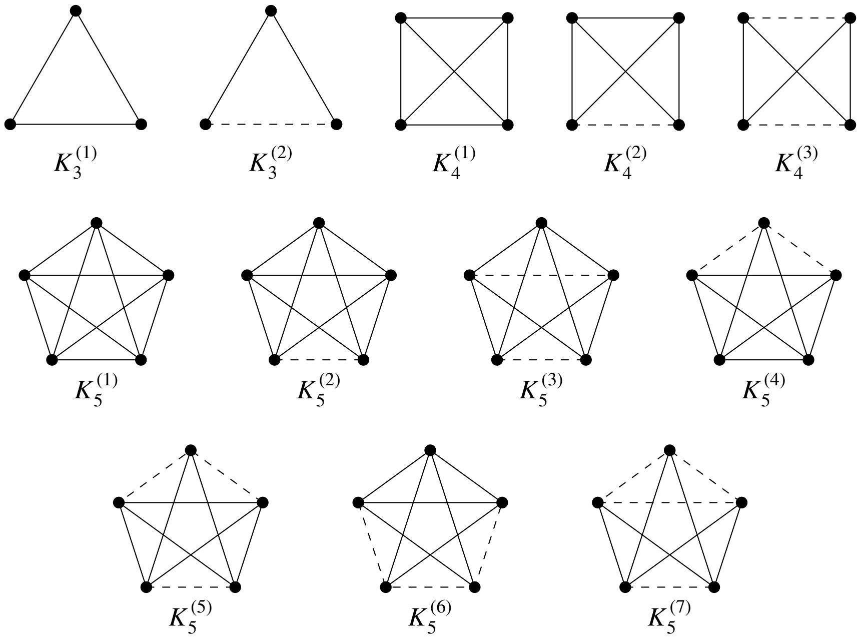


Figure 2. The switching classes of signed complete graphs.

We now review a few constructs on a signed graph $\Sigma = (V, E, \sigma)$ and describe our implementation. The *restriction* of Σ to an edge set $F \subseteq E$ is the signed graph $(V, F, \sigma|_F)$. For $e \in E$, we denote by $\Sigma - e$ (the *deletion* of e) the restriction of Σ to $E - \{e\}$. For $v \in V$, denote by $\Sigma - v$ the restriction of Σ to $E - F$, where F is the set of all edges incident to v. A component of the signed graph $\Sigma = (\Gamma, \sigma)$ is *balanced* if it contains no half-edges and each cycle has positive sign product.

Switching Σ by $s \in \{\pm\}^V$ results in the new signed graph (V, E, σ^s), where $\sigma^s_{vw} = s_v\, \sigma_{vw}\, s_w$. Switching does not alter balance, and any balanced signed graph can be obtained from switching an all-positive graph [Zaslavsky 1982b]. We also note that there is a natural bijection of proper colorings of Σ and a switched version of it, and this bijection preserves the number of proper k-colorings. Thus the chromatic polynomials of Σ are invariant under switching.

The *contraction* of Σ by $F \subseteq E$, denoted by Σ/F, is defined as follows [Zaslavsky 1982b]: switch Σ so that every balanced component of F is all positive, coalesce all nodes of each balanced component, and discard the remaining nodes and all edges in F; note that this may produce half-edges. If $F = \{e\}$ for a link e, Σ/e is obtained by switching Σ so that $\sigma(e) = +$ and then contracting e as in the case of unsigned graphs; that is, disregard e and identify its two endpoints. If e is a negative loop at v, then Σ/e has vertex set $V - \{v\}$ and edge set resulting from E

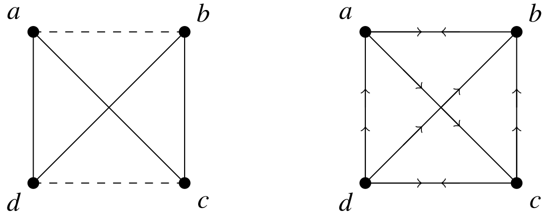
by deleting e and converting all edges incident with v to half-edges. The chromatic polynomial satisfies the deletion-contraction formula [Zaslavsky 1982a]

$$c_\Sigma(2k+1) = c_{\Sigma-e}(2k+1) - c_{\Sigma/e}(2k+1). \tag{1}$$

The zero-free chromatic polynomial $c^*_\Sigma(2k)$ satisfies the same identity provided that e is not a half-edge or negative loop. We will use (1) repeatedly in our computations.

We encode a signed graph Σ by its *incidence matrix* as follows: first *bidirect* Σ; i.e., give each edge an independent orientation at each endpoint (which we think of as an arrow pointing towards or away from the endpoint), such that a positive edge has one arrow pointing towards one and away from the other endpoint, and a negative edge has both arrows pointing either towards or away from the endpoints. The incidence matrix has rows indexed by vertices, columns indexed by edges, and entries equal to ± 1 according to whether the edge points towards or away from the vertex (and 0 otherwise). Since half-edges and negative loops have the same effect on the chromatic polynomial of Σ, we may assume that Σ has no half-edge. See Figure 3 for an example.

Deletion-contraction can be easily managed by incidence matrices: deletion of an edge simply means deletion of the corresponding column; contraction of a positive edge vw means replacing the rows corresponding to v and w by their sum and then deleting the column corresponding to the edge vw (it is sufficient to only consider contraction of positive edges, since we can always switch one of its endpoints if necessary, which means negating the corresponding row). Note that this process works for both links and half-edges. Note also that we will constantly look for multiple edges (with the same sign) and replace them with a single edge.



$$\begin{array}{ccccccc}
 & ab & ac & ad & bc & bd & cd \\
a & -1 & -1 & 1 & 0 & 0 & 0 \\
b & -1 & 0 & 0 & 1 & 1 & 0 \\
c & 0 & 1 & 0 & -1 & 0 & -1 \\
d & 0 & 0 & -1 & 0 & -1 & -1
\end{array}$$

Figure 3. $K_4^{(3)}$ with one of its bidirections and corresponding incidence matrix.

Thus we can keep track of incidence matrices as we recursively apply deletion-contraction, leading to empty signed graphs or signed graphs that only have half-edges; both have easy chromatic polynomials.

References

[Stein et al. 2012] W. A. Stein et al., "Sage mathematics software", 2012, available at http://www.sagemath.org. Version 5.1.

[Zaslavsky 1982a] T. Zaslavsky, "Signed graph coloring", *Discrete Math.* **39**:2 (1982), 215–228. MR 84h:05050a Zbl 0487.05027

[Zaslavsky 1982b] T. Zaslavsky, "Signed graphs", *Discrete Appl. Math.* **4**:1 (1982), 47–74. Erratum in **5**:2 (1983), p. 248. MR 84e:05095a Zbl 0476.05080

[Zaslavsky 1998–2012] T. Zaslavsky, "A mathematical bibliography of signed and gain graphs and allied areas", *Electron. J. Combin./Dyn. Surv.* **8** (1998–2012). MR 2000m:05001a Zbl 0898.05001

[Zaslavsky 2012] T. Zaslavsky, "Six signed Petersen graphs, and their automorphisms", *Discrete Math.* **312**:9 (2012), 1558–1583. MR 2899889 Zbl 1239.05086

Received: 2014-04-18 Revised: 2014-12-18 Accepted: 2015-01-13

mattbeck@sfsu.edu
Department of Mathematics, San Francisco State University,
San Francisco, CA 94132, United States

emeza2@lion.lmu.edu
Department of Mathematics, Loyola Marymount University,
Los Angeles, CA 90045, United States

nebryan@umich.edu
Department of Mathematics, Queens College, CUNY,
Flushing, NY 11367, United States

ashine@usc.edu
Department of Computer Science, University of Southern
California, Los Angeles, CA 90089, United States

myoung@iastate.edu
Department of Mathematics, Iowa State University,
Ames, IA 50011, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.833

Domino tilings of Aztec diamonds, Baxter permutations, and snow leopard permutations

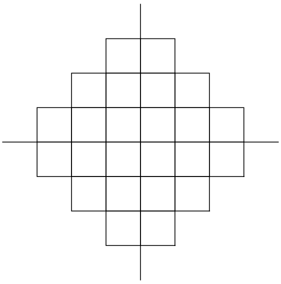
Benjamin Caffrey, Eric S. Egge, Gregory Michel, Kailee Rubin and Jonathan Ver Steegh

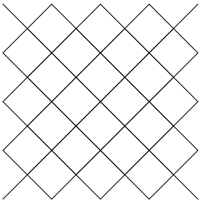
(Communicated by Arthur T. Benjamin)

In 1992, Elkies, Kuperberg, Larsen, and Propp introduced a bijection between domino tilings of Aztec diamonds and certain pairs of alternating-sign matrices whose sizes differ by one. In this paper we first study those smaller permutations which, when viewed as matrices, are paired with the matrices for doubly alternating Baxter permutations. We call these permutations snow leopard permutations, and we use a recursive decomposition to show they are counted by the Catalan numbers. This decomposition induces a natural map from Catalan paths to snow leopard permutations; we give a simple combinatorial description of the inverse of this map. Finally, we also give a set of transpositions which generates these permutations.

1. Introduction and background

An *Aztec diamond of order n* is a two-dimensional array of unit squares with $2i$ squares in rows $i \le n$ and $2(2n - i + 1)$ squares in rows $n < i \le 2n$, in which the squares are centered in each row. In the figure below (left) we have the Aztec diamond of order 3. We will be interested in the vertices of an Aztec diamond, which we prefer to arrange in rows and columns, so we will orient all of our Aztec diamonds as in the figure on the right.





MSC2010: 05A05, 05A15.

Keywords: domino tiling, Aztec diamond, Baxter permutation, alternating permutation, alternating-sign matrix, Catalan number.

Aztec diamonds can be tiled using 2×1 domino rectangles, which is to say they can be completely covered by disjoint dominoes whose union is the entire diamond. We call a tiling of an Aztec diamond with dominoes a TOAD for short.

In [Elkies et al. 1992], Elkies, Kuperberg, Larsen, and Propp describe how to construct, for each TOAD T of order n, a pair of matrices $\mathit{SASM}(T)$ and $\mathit{LASM}(T)$ of sizes $n \times n$ and $(n+1) \times (n+1)$, respectively. Each of these matrices is an *alternating-sign matrix* (ASM), which is a matrix with entries in $\{0, 1, -1\}$ whose nonzero entries in each row and in each column alternate in sign and sum to 1. (For an introduction to ASMs and a variety of related combinatorial objects, see [Robbins 1991; Bressoud 1999; Propp 2001].) To carry out this construction, first note that in Figure 1 the vertices that compose the tiled Aztec diamond fall naturally into two matrices: the red vertices form an $(n+1) \times (n+1)$ matrix while the blue vertices form an $n \times n$ matrix. We construct $\mathit{LASM}(T)$ on the red vertices by labeling each vertex of degree 4 with a 1, labeling each vertex of degree 3 with a 0, and labeling each vertex of degree 2 with a -1. We construct $\mathit{SASM}(T)$ on the blue vertices in the same way, except the degree 4 and degree 2 rules are reversed. Note that the TOAD T in Figure 1 has

$$\mathit{LASM}(T) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathit{SASM}(T) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Following [Elkies et al. 1992] and [Canary 2010], we say an $(n+1) \times (n+1)$ ASM A and an $n \times n$ ASM B are *compatible* whenever there is a TOAD T such that $A = \mathit{LASM}(T)$ and $B = \mathit{SASM}(T)$. Elkies et al. showed that an $(n+1) \times (n+1)$ ASM with k entries equal to -1 is compatible with 2^k $n \times n$ ASMs, while an $n \times n$ ASM with j entries equal to 1 is compatible with 2^j $(n+1) \times (n+1)$ ASMs. In general, then, the compatibility relation is not one-to-one. However, each $(n+1) \times (n+1)$ ASM with no -1 entries (that is, each $(n+1) \times (n+1)$ permutation matrix) is

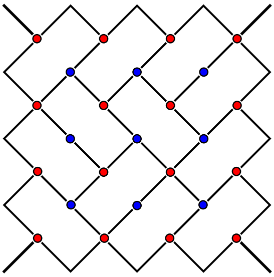


Figure 1. A domino tiling of the Aztec diamond of order 3.

compatible with exactly one $n \times n$ ASM. In this case, Canary [2010] gives an algorithm to construct the unique smaller ASM compatible with a given larger permutation matrix. (Asinowski [2014] gives a different formulation of the same algorithm, in which he first reconstructs the underlying TOAD.) To implement Canary's algorithm for an $(n+1) \times (n+1)$ permutation matrix A, first label the red vertices in a diagram for an Aztec diamond of the appropriate size with the entries of A. For each blue vertex, if the two red vertices immediately to the left, and all of the red vertices left of those, are labeled with 0, then label the blue vertex 0. Now repeat this process in each of the other three directions (up, right, and down). Canary shows that each row and column of blue vertices will now contain an odd number of unlabeled vertices, and there is a unique way to label these vertices with 1s and -1s to create an ASM.

Canary proves that the $n \times n$ ASM compatible with a given $(n+1) \times (n+1)$ permutation matrix A will also be a permutation matrix if and only if A is the matrix of a Baxter permutation. To understand the definition of a Baxter permutation, first note that we can interpret each permutation matrix A as the permutation π in one-line notation for which $A_{ij} = \delta_{j\pi(i)}$. That is, the 1 in the first row of A is in position $\pi(1)$, the 1 in the second row is in position $\pi(2)$, and in general the 1 in the jth row is in position $\pi(j)$. For example, if T is the TOAD in Figure 1, then the permutation for $LASM(T)$ is 4132 and the permutation for $SASM(T)$ is 312. We will often identify a permutation matrix with its corresponding permutation in one-line notation. With this convention, a *Baxter* permutation is a permutation that avoids 2–41–3 and 3–14–2. In other words, π is a Baxter permutation whenever there are no indices $i < j < j+1 < k$ such that $\pi(j+1) < \pi(i) < \pi(k) < \pi(j)$ (for 2–41–3) or $\pi(j) < \pi(k) < \pi(i) < \pi(j+1)$ (for 3–14–2). For example, 174962835 is not Baxter because the subsequence 4625 is an instance of 2–41–3. In contrast, 879164325 is Baxter because it contains no instances of 2–41–3 or 3–14–2. Note that the compatibility relation is still not one-to-one when we restrict it to Baxter permutations. For example, 12 is compatible with the Baxter permutations 123, 132, and 213. On the other hand, as suggested in [Asinowski et al. 2013], for every permutation π of length n which is compatible with a Baxter permutation of length $n+1$, the number of Baxter permutations of length $n+1$ compatible with π appears to be a product of Fibonacci numbers.

Baxter permutations first arose in connection with the question of whether two commuting continuous functions from the closed interval [0, 1] to itself must have a common fixed point [Baxter 1964; Boyce 1967]. Since their introduction they have been studied by many authors; some relevant references are [Chung et al. 1978; Mallows 1979; Cori et al. 1986; Dulucq and Guibert 1996; 1998; Guibert and Linusson 2000; Ouchterlony 2006; Ackerman et al. 2006; Asinowski et al. 2013].

Our work involves a particular class of Baxter permutations, which are known as doubly alternating Baxter permutations. We call a permutation π *alternating* whenever $\pi(i) < \pi(i+1)$ if i is odd and $\pi(i) > \pi(i+1)$ if i is even. That is, π is alternating whenever it begins with an ascent, and its ascents and descents alternate. A *doubly alternating* permutation is an alternating permutation whose inverse is also alternating, and we call permutations that are both doubly alternating and Baxter *doubly alternating Baxter permutations (DABPs).* Guibert and Linusson [2000] show that the Catalan number $C_n = 1/(n+1)\binom{2n}{n}$ counts both the DABPs of length $2n$ and the DABPs of length $2n+1$. The Catalan numbers are known to count many other combinatorial objects (see [Stanley 1999, Exercise 6.19] and [Stanley 2013]), including lattice paths from $(0, 0)$ to (n, n) using only north $(0, 1)$ and east $(1, 0)$ steps which do not pass below the line $y = x$; we call these paths *Catalan paths.* In addition to the explicit definition of C_n in terms of binomial coefficients, the Catalan numbers also satisfy the recurrence relation $C_n = \sum_{j=1}^{n} C_{j-1}C_{n-j}$ for $n \geq 0$, with initial condition $C_0 = 1$.

In this paper, we introduce the *snow leopard permutations (SLPs)*, which are the permutations that are compatible with the doubly alternating Baxter permutations. More formally, we write S_n to denote the set of permutations of length n, and we make the following definition.

Definition 1.1. We say a permutation $\pi \in S_n$ is a *snow leopard permutation* whenever there is a TOAD T of order n such that $LASM(T)$ is a DABP and $SASM(T) = \pi$.

In Section 2, we characterize these permutations recursively, and we use this recursive characterization to show that in this case the compatibility relation is one-to-one. This implies that the snow leopard permutations of length $2n$ are also counted by C_n, as are the snow leopard permutations of length $2n + 1$. Matching our recursive description of the snow leopard permutations with the first-return decomposition of a Catalan path gives us a recursively defined bijection from Catalan paths from $(0, 0)$ to (n, n) to snow leopard permutations of length $2n$. In Section 3 we give a simple combinatorial description of the inverse of this map. Finally, in Section 4 we describe how to generate all of the snow leopard permutations from the decreasing permutation with a specific set of transpositions.

2. Recursive decompositions of DABPs, TOADs, and snow leopard permutations

In this section we describe how to construct snow leopard permutations recursively, and we use our recursive decomposition to show that there are C_n snow leopard permutations of length $2n$, as well as C_n snow leopard permutations of length $2n+1$. Our snow leopard permutation decomposition is induced by similar decompositions of the associated TOADs and DABPs, so we first describe how to decompose these objects. We begin with a recursive decomposition of a DABP, for which it will be helpful to use several common operations on permutations.

Permutation tools. Throughout we write S_n to denote the set of all permutations of length n, and for any permutation π, we write $|\pi|$ to denote the length of π. The following four operations on permutations will be especially useful for us.

Definition 2.1. For any permutation $\pi \in S_n$, we write π^c to denote the *complement* of π, which is the permutation in S_n with

$$\pi^c(j) = n+1-\pi(j)$$

for all j, $1 \leq j \leq n$, and we write π^r to denote the *reverse* of π, which is the permutation in S_n with

$$\pi^r(j) = \pi(n+1-j)$$

for all j, $1 \leq j \leq n$. For any permutations $\pi \in S_n$ and $\sigma \in S_k$, we write $\pi \oplus \sigma$ to denote the permutation in S_{n+k} with

$$(\pi \oplus \sigma)(j) = \begin{cases} \pi(j) & \text{if } 1 \leq j \leq n, \\ n+\sigma(j-n) & \text{if } n < j \leq n+k \end{cases}$$

for all j, $1 \leq j \leq n$, and we write $\pi \ominus \sigma$ to denote the permutation in S_{n+k} with

$$(\pi \ominus \sigma)(j) = \begin{cases} k+\pi(j) & \text{if } 1 \leq j \leq n, \\ \sigma(j-n) & \text{if } n < j \leq n+k \end{cases}$$

for all j, $1 \leq j \leq n$.

Note that on matrices the complement is a reflection over a vertical line, while the reverse is a reflection over a horizontal line. In addition, one can also show that for any permutations π and σ, we have $(\pi \oplus \sigma)^{-1} = \pi^{-1} \oplus \sigma^{-1}$, $(\pi^r)^{-1} = (\pi^{-1})^c$, and $(\pi^c)^{-1} = (\pi^{-1})^r$. We sometimes write i to denote the inverse map on S_n; with this notation, our last two equations are equivalent to $i \circ r = c \circ i$ and $i \circ c = r \circ i$, respectively.

Example 2.2. If $\pi = 32154$ and $\sigma = 3124$ then $\pi^c = 34512$, $\sigma^r = 4213$, $\pi \oplus \sigma = 321548679$, and $\pi \ominus \sigma = 765983124$.

In some situations our permutations will naturally have length 0 or -1. To incorporate these cases into our results, we use the following notation.

Definition 2.3. We write $\varnothing$ to denote the empty permutation, which is the unique permutation of length 0, and we write @ to denote the *antipermutation*, which is the unique permutation of length -1. We have $@^c = @^r = @^{-1} = @$, and $1 \oplus @ = @ \oplus 1 = 1 \ominus @ = @ \ominus 1 = \varnothing$.

As we show next, the set of Baxter permutations is closed under $\oplus$, $\ominus$, taking complements, and taking the reverse of a permutation.

Lemma 2.4. *The following are equivalent for any permutation π.*

(i) *π is Baxter.*

(ii) *π^c is Baxter.*

(iii) *π^r is Baxter.*

(iv) *π^{-1} is Baxter.*

Proof. (i) $\Rightarrow$ (ii) If π^c contains a subsequence of type 2–41–3, then the corresponding subsequence of π will have type 3–14–2. Similarly, if π^c contains a subsequence of type 3–14–2 then the corresponding subsequence of π will have type 2–41–3. If π is Baxter then π avoids 2–41–3 and 3–14–2, so π^c avoids 3–14–2 and 2–41–3, which means π^c is Baxter.

(ii) $\Rightarrow$ (i) This is immediate from (i) $\Rightarrow$ (ii), since $(\pi^c)^c = \pi$.

(i) $\Leftrightarrow$ (iii) This is similar to the proof of (i) $\Leftrightarrow$ (ii).

(i) $\Leftrightarrow$ (iv) Since $(\pi^{-1})^{-1} = \pi$, it's sufficient to show that if π contains a subsequence of type 2–41–3 or a subsequence of type 3–14–2 then π^{-1} does, as well. With this in mind, suppose $abcd$ is a subsequence of π of type 2–41–3 for which $d - a$ is minimal. If $d = a + 1$ then the corresponding subsequence in π^{-1} has type 3–14–2. Otherwise, $a + 1$ is either to the left of b or to the right of c, since b and c are adjacent. If $a + 1$ is to the left of b, then we can replace a with $a + 1$, so $d - a$ was not minimal, which is a contradiction. On the other hand, if $a + 1$ is to the right of c then we can replace d with $a + 1$, so $d - a$ was not minimal in this case, either.

The proof that if π contains a subsequence of type 3–14–2 then π^{-1} contains a subsequence of type 2–41–3 or 3–14–2 is similar. □

Lemma 2.5. *The following are equivalent for permutations π and σ.*

(i) *π and σ are Baxter.*

(ii) *$\pi \oplus \sigma$ is Baxter.*

(iii) *$\pi \ominus \sigma$ is Baxter.*

Proof. (i) $\Rightarrow$ (ii) Suppose to the contrary that π and σ are Baxter permutations but $\pi \oplus \sigma$ is not Baxter. Call the first $|\pi|$ entries of $\pi \oplus \sigma$ the front of $\pi \oplus \sigma$, and call the last $|\sigma|$ entries the back. Note that every entry in the front is less than every entry in the back.

If $\pi \oplus \sigma$ contains a subsequence α of type 2–41–3, then α cannot be entirely contained in the front or in the back, since π and σ are Baxter. Therefore $\alpha(1)$ is in the front and $\alpha(4)$ is in the back. Now $\alpha(2)$ must be in the back, since it is greater than $\alpha(4)$, so $\alpha(3)$ must also be in the back. But this contradicts the fact that $\alpha(1) > \alpha(3)$.

If $\pi \oplus \sigma$ contains a subsequence α of type 3–14–2, then α cannot be entirely contained in the front or in the back, since π and σ are Baxter. But this contradicts the fact that $\alpha(1) > \alpha(4)$.

(ii) $\Rightarrow$ (i) If π or σ contains a subsequence of type 2–41–3 or 3–14–2 then so does $\pi \oplus \sigma$, and the result follows.

(i) $\Leftrightarrow$ (iii) This is similar to the proof of (i) $\Leftrightarrow$ (ii). □

Note that if π is alternating then π^c is not alternating in general, and π^r is alternating if and only if π has odd length. Similarly, if π and σ are alternating, then $\pi \oplus \sigma$ is not alternating in general, while $\pi \ominus \sigma$ is alternating if and only if π has even length. As a result, the set of DABPs is not closed under $\oplus$, $\ominus$, complements, or reverses.

The DABP decompositions. As we will see, snow leopard permutations inherit their recursive structure from DABPs, so our first goal is to describe how to decompose DABPs into smaller DABPs. Several of these results are not new, so we will refer to the work of others, especially [Dulucq and Guibert 1998] and [Ouchterlony 2006], as needed.

Lemma 2.6 [Ouchterlony 2006, Lemma 4.1(i)]. *If π is a DABP of odd length then* $\pi(1) = 1$.

Ouchterlony uses Lemma 2.6 to conclude that π is a DABP of length $2n+1$ if and only if $\pi = 1 \oplus (\sigma^r)^{-1}$ for some DABP σ of length $2n$ [Ouchterlony 2006, Corollary 4.2(i)], and that this correspondence is a bijection between the set of DABPs of length $2n+1$ and the set of DABPs of length $2n$. However, as we show next, more is true.

Proposition 2.7. *Suppose f is any of the functions r, c, $i \circ r$, and $i \circ c$ on permutations. For any nonnegative integer n and any $\pi \in S_{2n+1}$, π is a DABP if and only if there is a DABP $\sigma \in S_{2n}$ such that $\pi = 1 \oplus \sigma^f$. Moreover, for each f, this correspondence is a bijection between the set of DABPs π of length $2n+1$ and the set of DABPs σ of length $2n$.*

Proof. By [Ouchterlony 2006, Corollary 4.2(i)], the result holds for $f = i \circ r$. To prove the result for $f = c$, first note that σ is a DABP if and only if σ^{-1} is a DABP by Lemma 2.4. Now the result follows by replacing σ with σ^{-1} in [loc. cit., Corollary 4.2(i)] and using the fact that $i \circ r \circ i = c$.

The proofs when $f = r$ and $f = i \circ c$ are similar. □

With Proposition 2.7 in mind, we will focus our attention on DABPs of even length. In this case, Guibert and Linusson [2000] and Ouchterlony [2006] have found the following DABP decomposition.

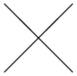


Figure 2. The TOAD of order 0.

Proposition 2.8 [Ouchterlony 2006, Corollary 4.2(ii)] and [Guibert and Linusson 2000, proof of Theorem 3]. *For any nonnegative integer n and any permutation $\pi \in S_{2n}$, π is a DABP if and only if there are DABPs π_1 and π_2 of even length such that $\pi = (1 \oplus (\pi_1^r)^{-1} \oplus 1) \ominus \pi_2$. Moreover, this correspondence is a bijection between the set of DABPs π of length $2n$ and the set of ordered pairs (π_1, π_2) of DABPs of lengths $2k$ and $2l$, where $n = k + l + 1$.*

As was the case for DABPs of odd length, more is true.

Proposition 2.9. *Suppose f is any of the functions r, c, $i \circ r$, and $i \circ c$ on permutations. For any nonnegative integer n and any permutation $\pi \in S_{2n}$, π is a DABP if and only if there are DABPs π_1 and π_2 of even length such that $\pi = (1 \oplus \pi_1^f \oplus 1) \ominus \pi_2$. Moreover, for each f, this correspondence is a bijection between the set of DABPs π of length $2n$ and the set of ordered pairs (π_1, π_2) of DABPs of lengths $2k$ and $2l$, where $n = k + l + 1$.*

Proof. This is similar to the proof of Proposition 2.7, using Proposition 2.8. □

The Aztec diamond decompositions. It is not difficult to show [Asinowski 2014; Canary 2010] that each Baxter permutation π of length $n+1$ determines a unique TOAD $\mathcal{T}(\pi)$ of order n, and that $\mathcal{T}$ and *LASM* are inverse bijections when *LASM* is restricted to those TOADS whose *LASM* is a Baxter permutation. Computing $\mathcal{T}(\pi)$ when π has length 2 or more is routine, but some care is required when π has length 0 or 1. In particular, $\mathcal{T}(1)$ is the TOAD of order 0, which we show in Figure 2. Going a bit smaller still, we write @ to denote the TOAD $\mathcal{T}(\varnothing)$, which has order -1. Since the Aztec diamond of order -1 has no edges at all, we can't even draw it, but it will still play a role in our snow leopard decomposition.

The fact that we have the maps $\mathcal{T}$ and *LASM* means our DABP decompositions induce similar TOAD decompositions. To describe these TOAD decompositions, it's useful to introduce several ways of transforming and combining TOADs.

Definition 2.10. For any TOAD T, we write T^c to denote the *complement* of T, which is the reflection of T over a vertical line, we write T^r to denote the *reverse* of T, which is the reflection of T over a horizontal line, and we write T^{-1} to denote the *inverse* of T, which is the reflection of T over a diagonal line from upper left to lower right.

As we did for permutations, we sometimes write i to denote the inverse map on TOADs.

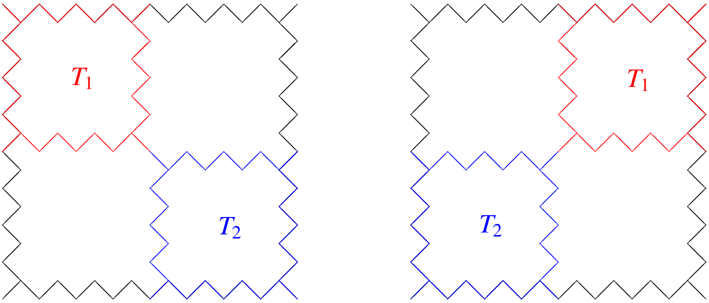


Figure 3. The construction of $T_1 \oplus T_2$ (left) and $T_1 \ominus T_2$ (right) from T_1 and T_2.

Definition 2.11. For any TOADs T_1 and T_2, we write $T_1 \oplus T_2$ to denote the TOAD we obtain by identifying the lower right vertex of T_1 with the upper left vertex of T_2, taking the smallest Aztec diamond D which contains both T_1 and T_2, and tiling the part of D outside of T_1 and T_2 with dominoes whose long sides are oriented from upper left to lower right. If T_1 has order n and T_2 has order k, then $T_1 \oplus T_2$ has order $n+k+1$.

In Figure 3 (left) we see how TOADs T_1 (in red) and T_2 (in blue) are combined to produce $T_1 \oplus T_2$. Note that the only way to tile the areas outside of T_1 and T_2 is to use dominoes whose long sides are oriented from upper left to lower right, as in the construction of $T_1 \oplus T_2$.

Definition 2.12. For any TOADs T_1 and T_2, we write $T_1 \ominus T_2$ to denote the TOAD we obtain by identifying the lower left vertex of T_1 with the upper right vertex of T_2, taking the smallest Aztec diamond D which contains both T_1 and T_2, and tiling the part of D outside of T_1 and T_2 with dominoes whose long sides are oriented from upper right to lower left. If T_1 has order n and T_2 has order k, then $T_1 \ominus T_2$ has order $n+k+1$.

In Figure 3 (right) we see how TOADs T_1 (in red) and T_2 (in blue) are combined to produce $T_1 \ominus T_2$. Note that the only way to tile the areas outside of T_1 and T_2 is to use dominoes whose long sides are oriented from upper right to lower left, as in the construction of $T_1 \ominus T_2$.

Our next result, which follows immediately from our definitions, justifies our multiple uses of the notations c, r, -1, $\oplus$, and $\ominus$.

Proposition 2.13. *For any Baxter permutations π and σ, the following hold.*

(i) $\mathcal{T}(\pi^c) = \mathcal{T}(\pi)^c$.

(ii) $\mathcal{T}(\pi^r) = \mathcal{T}(\pi)^r$.

(iii) $\mathcal{T}(\pi^{-1}) = \mathcal{T}(\pi)^{-1}$.

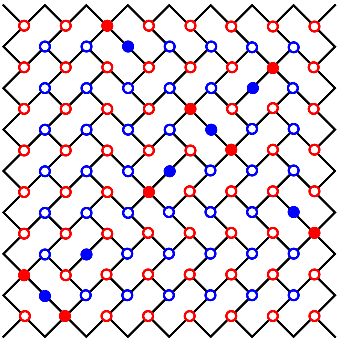


Figure 4. The DAAD corresponding to the DABP 37564812 and its compatible SLP 3654721.

(iv) $\mathcal{T}(\pi \oplus \sigma) = \mathcal{T}(\pi) \oplus \mathcal{T}(\sigma)$.

(v) $\mathcal{T}(\pi \ominus \sigma) = \mathcal{T}(\pi) \ominus \mathcal{T}(\sigma)$.

We now turn our attention to those TOADs which come from DABPs.

Definition 2.14. We call a TOAD T a *doubly alternating Aztec diamond (DAAD)* whenever $LASM(T)$ is a DABP. Note that a TOAD T is a DAAD if and only if there is a DABP π such that $\mathcal{T}(\pi) = T$. Indeed, $\pi = LASM(T)$.

In Figure 4 we have a DAAD with its DABP and its corresponding snow leopard permutation.

We saw in Proposition 2.7 that it's easy to construct DABPs of odd length from DABPs of even length. As we see next, this means it's easy to construct DAADs of even order from DAADs of odd order.

Proposition 2.15. *Suppose f is any of the functions r, c, $i \circ r$, and $i \circ c$ on DAADs. For any nonnegative integer n and any TOAD T of order $2n$, T is a DAAD if and only if there is a DAAD D of order $2n-1$ such that $T = \mathcal{T}(1) \oplus D^f$. Moreover, for each f this correspondence is a bijection between the set of DAADs of order $2n$ and the set of DAADs of order $2n-1$.*

Proof. ($\Rightarrow$) Since T is a DAAD of order $2n$, there is a DABP π of length $2n+1$ with $\mathcal{T}(\pi) = T$. By Proposition 2.7, there is a DABP σ of length $2n$ such that $\pi = 1 \oplus \sigma^f$. If we apply $\mathcal{T}$ to our expression for π and use Proposition 2.13 to simplify the result, we find $T = \mathcal{T}(1) \oplus \mathcal{T}(\sigma)^f$. Now the result follows, since $D = \mathcal{T}(\sigma)$ is a DAAD of order $2n-1$.

($\Leftarrow$) Since D is a DAAD of order $2n-1$, there is a DABP σ of length $2n$ such that $\mathcal{T}(\sigma) = D$. By Proposition 2.7, we have $\mathcal{T}(1 \oplus \sigma^f) = T$, so T is a DAAD.

The fact that this correspondence is a bijection follows from the last statement of Proposition 2.7 and the fact that $\mathcal{T}$ is a bijection. □

Proposition 2.15 says that we can understand all DAADs if we understand DAADs of odd order. With this in mind, we now describe how to decompose a DAAD of odd order into a combination of two smaller DAADs of odd order.

Theorem 2.16. *Suppose f is any of the functions r, c, $i \circ r$, or $i \circ c$ on TOADs. For any TOAD T of odd order, T is a DAAD if and only if there are DAADs T_1 and T_2 of odd order such that $T = (\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2$. Moreover, for each f, this correspondence is a bijection between the set of DAADs T of order $2n-1$ and the set of ordered pairs (T_1, T_2) of DAADs of orders $2k-1$ and $2l-1$, where $n = k+l+1$.*

Proof. ($\Rightarrow$) Since T is a DAAD or order $2n-1$, we know that $\pi = LASM(T)$ is a DABP of length $2n$ with $T = \mathcal{T}(\pi)$. By Proposition 2.9, there are DABPs π_1 and π_2 of lengths $2k$ and $2l$, respectively, such that $\pi = (1 \oplus \pi_1^f \oplus 1) \ominus \pi_2$ and $n = k+l+1$. If we apply $\mathcal{T}$ to our expression for π and use Proposition 2.13 to simplify the result, we find

$$T = \mathcal{T}(\pi) = \big(\mathcal{T}(1) \oplus \mathcal{T}(\pi_1)^f \oplus \mathcal{T}(1)\big) \ominus \mathcal{T}(\pi_2).$$

Now the result follows, since $T_1 = \mathcal{T}(\pi_1)$ and $T_2 = \mathcal{T}(\pi_2)$ are DAADs by definition.

($\Leftarrow$) Since T_1 and T_2 are DAADs, we know that $\pi_1 = LASM(T_1)$ and $\pi_2 = LASM(T_2)$ are DABPs of lengths k and l respectively such that $\mathcal{T}(\pi_1) = T_1$ and $\mathcal{T}(\pi_2) = T_2$. Moreover, $n = k+l+1$. By Proposition 2.9, the permutation $(1 \oplus \pi_1^f \oplus 1) \ominus \pi_2$ is also a DABP, so its image under $\mathcal{T}$ is a DAAD. But if we apply $\mathcal{T}$ to $(1 \oplus \pi_1^f \oplus 1) \ominus \pi_2$ and use Proposition 2.13 to simplify the result, we find that

$$\mathcal{T}((1 \oplus \pi_1^f \oplus 1) \ominus \pi_2) = (\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2.$$

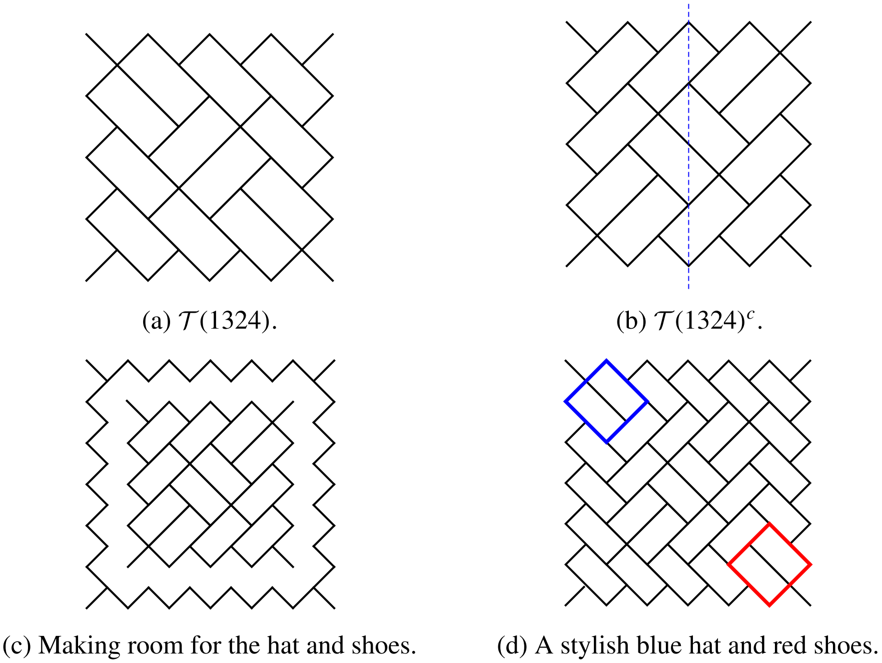
Therefore $(\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2$ is a DAAD.

The fact that this correspondence is a bijection follows from the last statement of Proposition 2.9 and the fact that $\mathcal{T}$ is a bijection. □

When we consider how our DAAD decomposition gives us a decomposition of the associated snow leopard permutation, we will be especially interested in pairs of dominoes that share a long side. With this in mind, we sometimes think of the process of building $\mathcal{T}(1) \oplus T \oplus \mathcal{T}(1)$ from a TOAD T in terms of adding a "hat" and pair of "shoes" to T. In Figure 5 we add a hat (in blue) and shoes (in *Wizard of Oz* ruby red) to $\mathcal{T}(1324)^c$.

When we construct $(\mathcal{T}(1) \oplus T_1 \oplus \mathcal{T}(1)) \ominus T_2$ from $\mathcal{T}(1) \oplus T_1 \oplus \mathcal{T}(1)$ and T_2, we add one more pair of dominoes which are adjacent along long sides; we call this pair the "connector". In Figure 6(d) we outline the connector in red.

The snow leopard permutation decompositions. In the Introduction we described the function *SASM*, which maps DAADs of order n to snow leopard permutations


(a) $\mathcal{T}(1324)$.

(b) $\mathcal{T}(1324)^c$.

(c) Making room for the hat and shoes.

(d) A stylish blue hat and red shoes.

Figure 5. An illustration of the computation of $\mathcal{T}(1) \oplus \mathcal{T}(1324)^c \oplus \mathcal{T}(1)$, also known as the "hat and shoes" process.

of length n. In this section we use *SASM* and our DAAD decomposition to obtain our snow leopard permutation decomposition. To make this easier, we first describe a simple relationship between certain domino configurations in a DAAD T and the 1s in the matrix for $SASM(T)$.

Definition 2.17. A *block* in a TOAD T is a pair of two dominoes in T which are adjacent along a long edge, forming a 2-by-2 box.

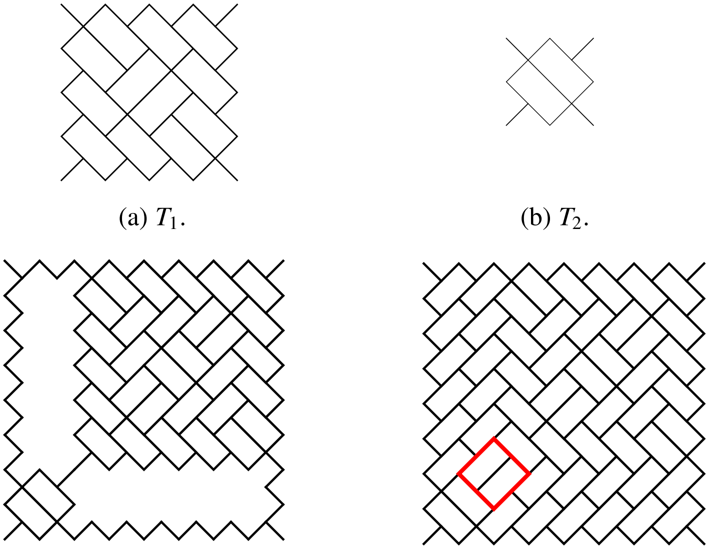
The DAAD shown in Figure 4 contains 7 blocks.

Lemma 2.18. *The vertices in a DAAD T which correspond to the* 1*s in $SASM(T)$ are exactly those vertices in the center of a block. As a result, the blocks of a DAAD are in bijection with the* 1*s in its SASM.*

Proof. Let T be a DAAD of order n that contains a block B. By Canary's algorithm, this point may correspond to a 1 in $SASM(T)$ or a -1 in $LASM(T)$. However, because $LASM(T)$ is a permutation, it cannot contain a -1. Thus, a block must correspond to a 1 in $SASM(T)$.

Conversely, a 1 in $SASM(T)$ must label a vertex of degree 2, which creates a block in T. □

Next we describe how the map *SASM* interacts with our operations on TOADs.



(a) T_1.

(b) T_2.

(c) $\mathcal{T}(1) \oplus T_1^c \oplus \mathcal{T}(1)$ and T_2 in a a larger diamond.

(d) Putting the rest of the dominoes in, with the connector in red.

Figure 6. An illustration of the composition of DAADs T_1 and T_2, using the complement map. We outline the connector in red.

Proposition 2.19. *For any TOADs T_1 and T_2, the following hold*:

(i) $SASM(T_1^c) = SASM(T_1)^c$.

(ii) $SASM(T_1^r) = SASM(T_1)^r$.

(iii) $SASM(T_1^{-1}) = SASM(T_1)^{-1}$.

(iv) $SASM(T_1 \oplus T_2) = SASM(T_1) \oplus 1 \oplus SASM(T_2)$.

(v) $SASM(T_1 \ominus T_2) = SASM(T_1) \ominus 1 \ominus SASM(T_2)$.

Proof. (i), (ii), (iii) These are clear from Lemma 2.18 and our construction of $SASM$, since each of c, r, and i is a reflection over a particular line.

(iv) First observe that if T_1 (resp. T_2) is the TOAD of order -1 then $T_1 \oplus T_2$ is equal to T_1 (resp. T_2). But in this case $SASM(T_1)$ (resp. $SASM(T_2)$) is the antipermutation @, and the result holds.

Now suppose T_1 and T_2 have nonnegative orders. Then in the construction of $T_1 \oplus T_2$ we create one block which is not in T_1 or T_2, where the lower right edge of T_1 meets the upper left edge of T_2. Now the result follows from Lemma 2.18.

(v) This is similar to the proof of (iv). □

We can now describe our snow leopard permutation decomposition.

Theorem 2.20. *Suppose f is any of the functions r, c, $i \circ r$, or $i \circ c$ on permutations. For any permutation π of odd length, π is a snow leopard permutation if and only if there are snow leopard permutations π_1 and π_2 of odd length such that $\pi = (1 \oplus \pi_1^f \oplus 1) \ominus 1 \ominus \pi_2$. Moreover, for each f, this correspondence is a bijection between the set of snow leopard permutations π of length $2n-1$ and the set of ordered pairs (π_1, π_2) of snow leopard permutations of lengths $2k-1$ and $2l-1$, where $n = k+l+1$.*

Proof. ($\Rightarrow$) If π is a snow leopard permutation of length $2n-1$, then by definition there is a DAAD T of order $2n-1$ such that $SASM(T) = \pi$. By Theorem 2.16, there are DAADs T_1 and T_2 of orders $2k-1$ and $2l-1$, where $n = k+l+1$, such that $T = (\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2$. Using Proposition 2.19, we find

$$\begin{aligned}
\pi &= SASM(T)\\
&= SASM\big((\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2\big)\\
&= \big(SASM(\mathcal{T}(1)) \oplus 1 \oplus SASM(T_1)^f \oplus 1 \oplus SASM(\mathcal{T}(1))\big) \ominus 1 \ominus SASM(T_2)\\
&= (1 \oplus SASM(T_1)^f \oplus 1) \ominus 1 \ominus SASM(T_2),
\end{aligned}$$

where the last step follows from the fact that $SASM(\mathcal{T}(1)) = \varnothing$. Now the result follows, since $\pi_1 = SASM(T_1)$ is a snow leopard permutation of length $2k-1$ and $\pi_2 = SASM(T_2)$ is snow leopard permutation of length $2l-1$, where $n = k+l+1$.

($\Leftarrow$) If π_1 and π_2 are snow leopard permutations of lengths $2k-1$ and $2l-1$, respectively, where $n = k+l+1$, then by definition there are DAADs T_1 and T_2 of orders $2k-1$ and $2l-1$, respectively, such that $\pi_1 = SASM(T_1)$ and $\pi_2 = SASM(T_2)$. By Theorem 2.16, we know that $(\mathcal{T}(1) \oplus T_1^f \oplus \mathcal{T}(1)) \ominus T_2$ is a DAAD of order $2n-1$. But if we apply $SASM$ to this DAAD and use Proposition 2.19 as in the proof of the other direction, we find $(1 \oplus \pi_1^f \oplus 1) \ominus 1 \ominus \pi_2$ is a snow leopard permutation of length $2n-1$.

To see that the map $(\pi_1, \pi_2) \mapsto (1 \oplus \pi_1^r \oplus 1) \ominus 1 \ominus \pi_2$ is a bijection, first note that it is onto the set of snow leopard permutations by the first part of the theorem. To see it is one-to-one, suppose there are ordered pairs (π_1, π_2) and (σ_1, σ_2) of snow leopard permutations such that $(1 \oplus \pi_1^f \oplus 1) \ominus 1 \ominus \pi_2 = (1 \oplus \sigma_1^f \oplus 1) \ominus 1 \ominus \sigma_2$, and let π denote this common permutation. Then the hat (the second 1 in $1 \oplus \pi_1^f \oplus 1$ and $1 \oplus \sigma_1^f \oplus 1$) corresponds to the largest entry in π. Therefore π_1^f is a shift of the entries between the first entry of π and the largest entry of π, as is σ_1^f, so $\pi_1^f = \sigma_1^f$. But f is invertible, so $\pi_1 = \sigma_1$. Similarly, π_2 and σ_2 are both equal to the sequence of entries of π to the right of the largest entry of π, so $\pi_2 = \sigma_2$. □

It's worth noting that in small cases the permutation $(1 \oplus \pi_1^f \oplus 1) \ominus 1 \ominus \pi_2$ is not as long as it looks. For example, the antipermutation @ of length -1 is a snow leopard permutation corresponding to the TOAD of order -1. As a result, the snow leopard permutation 1 corresponds to the ordered pair $(@, @)$, since $1 = (1 \oplus @ \oplus 1) \ominus 1 \ominus @$. Similarly, for any snow leopard permutation π of odd length, $1 \oplus \pi \oplus 1$ and $1 \ominus 1 \ominus \pi$ are also snow leopard permutations of odd length, corresponding to the ordered pairs $(\pi, @)$ and $(@, \pi)$, respectively.

We can now use Theorem 2.20 to count the snow leopard permutations of each length.

Corollary 2.21. *For each $n \geq 0$, the number of snow leopard permutations of length $2n-1$ is C_n.*

Proof. For each $n \geq 0$, let a_n be the number of snow leopard permutations of length $2n - 1$. There is just one snow leopard permutation of length -1, so $a_0 = 1 = C_0$ and the result holds for $n = 0$. Now fix $n \geq 1$ and suppose by induction that $a_j = C_j$ for all j, $0 \leq j \leq n-1$. By Theorem 2.20 and our induction hypothesis, we have

$$a_n = \sum_{j=0}^{n-1} a_j a_{n-1-j} = \sum_{j=0}^{n-1} C_j C_{n-1-j} = \sum_{j=1}^{n} C_{j-1} C_{n-j} = C_n. \qquad \square$$

We can also use Theorem 2.20 and Proposition 2.7 to count the snow leopard permutations of even length.

Proposition 2.22. *Suppose f is any of the functions r, c, $i \circ r$, or $i \circ c$ on permutations. Then for any $n \geq 0$, the map $\pi \mapsto 1 \oplus \pi^f$ is a bijection between the set of snow leopard permutations of length $2n - 1$ and the set of snow leopard permutations of length $2n$.*

Proof. We first show that π is a snow leopard permutation of length $2n-1$ if and only if $1 \oplus \pi^f$ is a snow leopard permutation of length $2n$.

If π is a snow leopard permutation of length $2n - 1$, then by definition there is a DAAD T of order $2n - 1$ such that $SASM(T) = \pi$. By Proposition 2.15, the TOAD $\mathcal{T}(1) \oplus D^f$ is a DAAD of order $2n$. Now by Proposition 2.19, we have $SASM(\mathcal{T}(1) \oplus D^f) = 1 \oplus \pi^f$, since $SASM(\mathcal{T}(1)) = \varnothing$. Therefore $1 \oplus \pi^f$ is a snow leopard permutation of length $2n$.

Conversely, if $1 \oplus \pi^f$ is a snow leopard permutation of length $2n$, then by definition there is a DAAD T of order $2n$ such that $SASM(T) = 1 \oplus \pi^f$. Now by Proposition 2.15, there is a DAAD D of order $2n - 1$ such that $T = \mathcal{T}(1) \oplus D^f$, and by Proposition 2.19, we have $SASM(T) = 1 \oplus SASM(D)^f$. Since π^f can be obtained from $1 \oplus \pi^f$ and f is invertible, we must have $\pi = SASM(D)$, so π is a snow leopard permutation.

Finally, it is routine to check that the map $\pi \mapsto 1 \oplus \pi^f$ is a bijection between S_{2n-1} and the set of permutations in S_{2n} whose first entry is 1, so the restriction of this map to the set of snow leopard permutations of length $2n-1$ must also be a bijection. $\square$

Corollary 2.23. *For each $n \geq 0$, the compatibility correspondence is a bijection between the set of DABPs of length n and the set of snow leopard permutations of length $n-1$.*

Proof. By definition the compatibility correspondence maps DABPs of length n onto snow leopard permutations of length $n-1$. Since each of these sets has the same number of elements, this correspondence must be a bijection. □

Theorem 2.20 also gives us useful structural information about snow leopard permutations. For instance, we have the following result concerning the parities of the entries of a snow leopard permutation.

Corollary 2.24. *Snow leopard permutations preserve parity. That is, if π is a snow leopard permutation of length n, then for all j with $1 \leq j \leq n$, the entry $\pi(j)$ is even if and only if j is even.*

Proof. We first consider the case in which n is odd.

The result is vacuously true for $\pi = @$, and trivial for $\pi = 1$, so suppose by induction that $n \geq 0$ is odd and the result holds for all snow leopard permutations of odd length less than n.

In general, if σ is a permutation of odd length which preserves parity, then σ^c, $1 \oplus \sigma$, and $1 \oplus \sigma \oplus 1$ also preserve parity. Similarly, if σ is a parity-preserving permutation of odd length then $1 \ominus \sigma$ is a parity-reversing permutation. Finally, if σ_1 is a parity-preserving permutation of odd length and σ_2 is a parity-reversing permutation of even length, then $\sigma_1 \ominus \sigma_2$ is a parity-preserving permutation.

By Theorem 2.20, if π is a snow leopard permutation of odd length then there are snow leopard permutations π_1 and π_2 of odd length such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. By induction and our observations above, $1 \oplus \pi_1^c \oplus 1$ is a parity-preserving permutation of odd length and $1 \ominus \pi_2$ is a parity-reversing permutation of even length, so π preserves parity.

Now suppose π is a snow leopard permutation of even length. By Proposition 2.22, we have $\pi = 1 \oplus \sigma^c$ for some snow leopard permutation σ of odd length. By our observations above, σ^c preserves parity, so $\pi = 1 \oplus \sigma^c$ also preserves parity. □

Theorem 2.20 also gives us pattern-avoidance properties of snow leopard permutations. In particular, we can use it to show that snow leopard permutations are *anti-Baxter*, which means they avoid 2–14–3 and 3–41–2.

Corollary 2.25. *If π is a snow leopard permutation then π avoids* 2–14–3 *and* 3–41–2.

Proof. We first consider the case in which $|\pi| = n$ is odd.

The result is clear for $\pi = @$, $\pi = 1$, $\pi = 123$, and $\pi = 321$, so suppose by induction that $n \geq 0$ is odd and the result holds for all snow leopard permutations of odd length less than n. By Theorem 2.20, if π is a snow leopard permutation of odd length then there are snow leopard permutations π_1 and π_2 of odd length such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. For convenience, we call the entries of π corresponding to $1 \oplus \pi_1^c \oplus 1$ the *front* of π, and we call the remaining entries of π the *back* of π. Note that every entry in the front of π is greater than every entry in the back of π.

Now suppose π contains a subsequence $abcd$ of type 2–14–3. If a is in the front of π, then d is also in the front of π, since $d > a$. Moreover, a cannot be the first entry of the front of π and d cannot be the last, since the first and last entries are the smallest and largest entries of the front of π, and we have $b < a$ and $c > d$. Therefore our subsequence is entirely contained in the entries of π corresponding to π_1^c, and the corresponding subsequence of π_1 has type 3–41–2. This contradicts our induction hypothesis.

On the other hand, if a is not in the front of π then every entry of our subsequence is in the back of π. The first entry of the back of π is the largest, but $c > a$, so in fact our subsequence is contained in π_2, which contradicts our induction hypothesis.

The proof that π has no subsequence of type 3–41–2 is similar.

Now suppose π is a snow leopard permutation of even length. By Proposition 2.22, we have $\pi = 1 \oplus \sigma^c$ for some snow leopard permutation σ of odd length. Arguing as above, if π has a subsequence of type 2–14–3 (resp. 3–41–2) then σ has a subsequence of type 3–41–2 (resp. 2–14–3), so the result follows by induction. □

One can show that this result holds more generally: if π is a Baxter permutation of length $n+1$ and σ is a compatible permutation of length n, then σ is anti-Baxter [Asinowski et al. 2013].

3. A bijection from snow leopard permutations to Catalan paths

Like the snow leopard permutations, Catalan paths have a natural recursive decomposition. In particular, every nonempty Catalan path with $2n$ steps has the form Np_1Ep_2, where p_1 and p_2 are Catalan paths with $2k$ and $2l$ steps, respectively, and $n = k+l-1$. In fact, this decomposition gives a bijection between the set of Catalan paths p with $2n$ steps and ordered pairs (p_1, p_2) of Catalan paths with $2k$ and $2l$ steps, where $n = k + l - 1$. Matching this decomposition with our snow leopard permutation decomposition gives us a natural bijection from the set of Catalan paths with $2n$ steps to the set of snow leopard permutations of length $2n - 1$.

Proposition 3.1. *Suppose f is any of the functions r, c, $i \circ r$, and $i \circ c$. Then for each nonnegative integer n, there is a unique bijection Γ_f from the set of Catalan paths with $2n$ steps to the set of snow leopard permutations of length $2n - 1$ such that $\Gamma_f(\varnothing) = @$ and*

$$\Gamma_f(Np_1Ep_2) = \big(1 \oplus \Gamma_f(p_1)^f \oplus 1\big) \ominus 1 \ominus \Gamma_f(p_2)$$

for any Catalan paths p_1 and p_2.

p	$\Gamma_c(p)$
$\varnothing$	@
NE	1
$NNEE$	123
$NENE$	321
$NNNEEE$	14325
$NNENEE$	12345
$NNEENE$	34521
$NENNEE$	54123
$NENENE$	54321

p	$\Gamma_c(p)$
$NNNNEEEE$	1634527
$NNNENEEE$	1654327
$NNNEENEE$	1432567
$NNNEEENE$	3654721
$NNENNEEE$	1236547
$NNENENEE$	1234567
$NNENEENE$	3456721
$NNEENENE$	5674321
$NNEENNEE$	5674123
$NENNNEEE$	7614325
$NENNENEE$	7612345
$NENNEENE$	7634521
$NENENNEE$	7654123
$NENENENE$	7654321

Table 1. Values of $\Gamma_c(p)$ for short Catalan paths p.

Proof. Since each nonempty Catalan path can be written uniquely in the form Np_1Ep_2, where p_1 and p_2 are Catalan paths, Γ_f is well-defined and unique.

To show that $\Gamma_f(p)$ is a snow leopard permutation for every Catalan path p, first note that this is true for $p = \varnothing$ and $p = NE$. Now suppose by induction that p is a Catalan path with at least 4 steps, and that the result holds for all Catalan paths with fewer steps. Then there are unique Catalan paths p_1 and p_2 such that $p = Np_1Ep_2$, and by definition we have $\Gamma_f(p) = \big(1 \oplus \Gamma_f(p_1)^f \oplus 1\big) \ominus 1 \ominus \Gamma_f(p_2)$. By induction $\Gamma_f(p_1)$ and $\Gamma_f(p_2)$ are snow leopard permutations, so by Theorem 2.20 we see that $\Gamma_f(p)$ is also a snow leopard permutation.

To show that Γ_f is onto, first note that this is true for $n = 0$ and $n = 1$, so fix $n \geq 2$ and suppose by induction that the result holds for all smaller values of n. If π is a snow leopard permutation of length $2n - 1$, then by Theorem 2.20 there are shorter snow leopard permutations π_1 and π_2 of odd length such that $\pi = (1 \oplus \pi_1^f \oplus 1) \ominus 1 \ominus \pi_2$. By induction there are Catalan paths p_1 and p_2 such that $\Gamma_f(p_1) = \pi_1$ and $\Gamma_f(p_2) = \pi_2$, and by the definition of Γ_f, we have $\Gamma_f(Np_1Ep_2) = \pi$.

Since the set of Catalan paths with $2n$ steps and the set of snow leopard permutations of length $2n - 1$ are equinumerous by Corollary 2.21, the map Γ_f must be a bijection. □

Although all four maps Γ_f are bijections, we will be particularly interested in Γ_c. In Table 1 we have the values of Γ_c for all Catalan paths with 8 or fewer steps.

π	$\kappa(\pi)$
@	$\varnothing$
1	*NE*
123	*NNEE*
321	*NENE*
14325	*NNNEEE*
12345	*NNENEE*
34521	*NNEENE*
54123	*NENNEE*
54321	*NENENE*

π	$\kappa(\pi)$
1634527	*NNNNEEEE*
1654327	*NNNENEEE*
1432567	*NNNEENEE*
3654721	*NNNEEENE*
1236547	*NNENNEEE*
1234567	*NNENENEE*
3456721	*NNENEENE*
5674321	*NNEENENE*
5674123	*NNEENNEE*
7614325	*NENNNEEE*
7612345	*NENNENEE*
7634521	*NENNEENE*
7654123	*NENENNEE*
7654321	*NENENENE*

Table 2. Values of $\kappa(\pi)$ for short snow leopard permutations π.

While it is not obvious from these data, it turns out that Γ_c^{-1} has a simple, direct description in terms of ascents and descents.

Definition 3.2. For any snow leopard permutation π of length $2n-1$, we write $\kappa(\pi)$ to denote the lattice path with $2n$ steps whose i-th step $\kappa(\pi)_i$ is given by

$$\kappa(\pi)_i = \begin{cases} N & \begin{array}{l} \pi(i) < \pi(i+1) \text{ and } i \text{ is odd} \\ \text{or} \\ \pi(i) > \pi(i+1) \text{ and } i \text{ is even,} \end{array} \\ E & \begin{array}{l} \pi(i) < \pi(i+1) \text{ and } i \text{ is even} \\ \text{or} \\ \pi(i) > \pi(i+1) \text{ and } i \text{ is odd} \end{array} \end{cases}$$

for $0 \le i \le 2n-1$. By convention, we treat the empty entries $\pi(0)$ and $\pi(2n)$ as $2n$ and 0, respectively.

Example 3.3. Consider the permutation $\pi = 789634521$, which has ascent/descent sequence $DAADDAADDD$. Thus we have $\kappa(\pi) = NNEENNEENE$.

In Table 2 we have the values of $\kappa(\pi)$ for all snow leopard permutations π of length 7 or less.

It is not immediately obvious that κ maps every snow leopard permutation to a Catalan path, so we prove this next.

Proposition 3.4. *Suppose π is a snow leopard permutation of length $2n-1$. Then $\kappa(\pi)$ is a Catalan path of length $2n$.*

Proof. It is routine to check this when π has length 3 or less, since $\kappa(@) = \varnothing$, $\kappa(1) = NE$, $\kappa(123) = NNEE$, and $\kappa(321) = NENE$. Now suppose the result holds for all snow leopard permutations of odd length less than $2n-1$, where $2n-1 \geq 5$, and that π is a snow leopard permutation of length $2n-1$. By Theorem 2.20, there are snow leopard permutations π_1 and π_2 of lengths $2k-1$ and $2l-1$, respectively, such that $n = k+l+1$ and $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. We now consider three cases.

Case One: If $\pi_1 = @$ then $\pi = 1 \ominus 1 \ominus \pi_2$. In this case the ascent/descent sequence for π consists of two descents, followed by the ascent/descent sequence for π_2. By the definition of κ, this means $\kappa(\pi) = NE\kappa(\pi_2)$. Since $\kappa(\pi_2)$ is a Catalan path by induction, so is $\kappa(\pi)$.

Case Two: If $\pi_2 = @$ then $\pi = 1 \oplus \pi_1^c \oplus 1$. Since the complement operation on permutations turns ascents into descents and vice versa, the ascent/descent sequence for π consists of a descent, followed by the complement of the ascent/descent sequence for π_1, followed by a descent. By the definition of κ, this means $\kappa(\pi) = N\kappa(\pi_1)E$. Since $\kappa(\pi_1)$ is a Catalan path by induction, so is $\kappa(\pi)$.

Case Three: Suppose $\pi_1 \neq @$ and $\pi_2 \neq @$. Reasoning as in the previous cases, we find that the ascent/descent sequence for π consists of a descent, followed by the complement of the ascent/descent sequence for π_1, followed by an E, followed by the ascent/descent sequence for π_2. By the definition of κ, $\kappa(\pi) = N\kappa(\pi_1)E\kappa(\pi_2)$. Since $\kappa(\pi_1)$ and $\kappa(\pi_2)$ are Catalan paths by induction, so is $\kappa(\pi)$. □

The data in Tables 1 and 2, along with a close examination of the proof of Proposition 3.4, suggest that Γ_c and κ are inverses of one another; we prove this next.

Theorem 3.5. *Γ_c and κ are inverse functions.*

Proof. By Proposition 3.1 we know that Γ_c maps Catalan paths with $2n$ steps to snow leopard permutations of length $2n-1$, and by Proposition 3.4, the function κ maps snow leopard permutations of length $2n-1$ to Catalan paths with $2n$ steps. Since Γ_c is invertible, it's sufficient to show that $\Gamma_c(\kappa(\pi)) = \pi$ for every snow leopard permutation π.

The result is routine to check for $\pi = @$ and $\pi = 1$, so suppose π has length $2n-1 > 1$ and the result holds for all shorter snow leopard permutations. By Theorem 2.20, there are snow leopard permutations π_1 and π_2 such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. Reasoning as in the proof of Proposition 3.4, we see that $\kappa(\pi) = N\kappa(\pi_1)E\kappa(\pi_2)$. Now by the definition of Γ_c and our induction hypothesis, we have

$$\begin{aligned}\Gamma_c(\kappa(\pi)) &= \Gamma_c(N\kappa(\pi_1)E\kappa(\pi_2)) \\ &= N(\Gamma_c(\kappa(\pi_1)))^c E\Gamma_c(\kappa(\pi_2)) \\ &= N\pi_1^c E\pi_2 \\ &= \pi.\end{aligned}$$

□

4. Using transpositions to generate snow leopard permutations

It is well known that every permutation is a product of adjacent transpositions, so the adjacent transpositions generate S_n. In this section we introduce a simple set of transpositions, and we show that the snow leopard permutations of odd length are exactly the permutations one can construct from the decreasing permutation using sequences of our transpositions. We begin with the transpositions themselves.

Definition 4.1. Suppose that π is a permutation with consecutive entries $\pi(i)$, $\pi(i+1), \ldots, \pi(j)$.

(1) If $\pi(i)$ and $\pi(j)$ are odd and either $\pi(i-1), \pi(i), \ldots, \pi(j), \pi(j+1)$ or $\pi(i-1), \pi(j), \ldots, \pi(i), \pi(j+1)$ is a decreasing sequence of consecutive integers, and σ is the permutation we obtain from π by interchanging $\pi(i)$ and $\pi(j)$, then we say π and σ are *related by* τ_1.

(2) If $\pi(i)$ and $\pi(j)$ are even and either $\pi(i-1), \pi(i), \ldots, \pi(j), \pi(j+1)$ or $\pi(i-1), \pi(j), \ldots, \pi(i), \pi(j+1)$ is an increasing sequence of consecutive integers, and σ is the permutation we obtain from π by interchanging $\pi(i)$ and $\pi(j)$, then we say π and σ are *related by* τ_2.

By convention, if $\pi(i)$ or $\pi(j)$ occurs at either end of π, then we waive any requirement for the behavior of π beyond that point.

Example 4.2. The permutations $\pi = 983654721$ and $\sigma = 983456721$ are related by τ_2, since 36547 can be replaced with 34567.

Example 4.3. The permutations $\pi = 567894321$ and $\sigma = 567894123$ are related by τ_1, since 4321 can be replaced with 4123.

In Figure 7 we have graphs showing how the snow leopard permutations of lengths 3, 5, and 7 are related to one another by τ_1 and τ_2.

Although we don't do it here, one can study the parity of the number of inversions in a snow leopard permutation of odd length to show that these graphs are always bipartite.

As we show next, snow leopard permutations are only related to other snow leopard permutations by τ_1 and τ_2. We begin with a lemma concerning snow leopard permutations which begin with a decreasing sequence of consecutive integers.

Lemma 4.4. *If π is a snow leopard permutation of odd length, and there is a permutation σ of odd length with $\pi = 1 \ominus 1 \ominus \cdots \ominus 1 \ominus \sigma$, then σ is a snow leopard permutation.*

Proof. We argue by induction on $|\pi| - |\sigma|$.

If $|\pi| = |\sigma|$ then $\pi = \sigma$, and the result is clear. If $|\pi| - |\sigma| = 2$ then $\pi = 1 \ominus 1 \ominus \sigma = (1 \oplus @ \oplus 1) \ominus 1 \ominus \sigma$ must be the snow leopard decomposition of π guaranteed by Theorem 2.20, so σ is a snow leopard permutation.

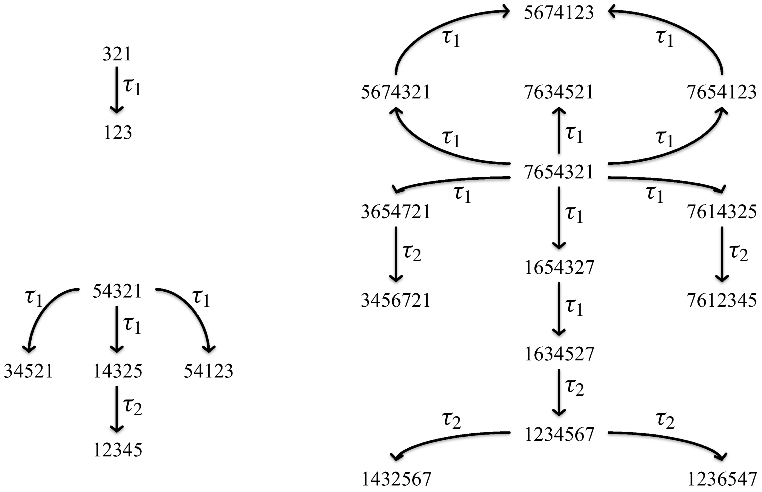


Figure 7. Graphs showing how the snow leopard permutations of lengths 3, 5, and 7 are related by τ_1 and τ_2.

Now suppose $|\pi| - |\sigma| \geq 4$. By Theorem 2.20, there are snow leopard permutations π_1 and π_2 such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. But π begins with its largest element, so we must have $\pi_1 = @$ and $\pi = 1 \ominus 1 \ominus \pi_2$. Therefore π_2 has the same form as π, but with two fewer 1s, so by induction σ is a snow leopard permutation. □

Theorem 4.5. *Suppose π is a snow leopard permutation of odd length and σ is a permutation.*

(i) *If π and σ are related by τ_1, then σ is a snow leopard permutation.*

(ii) *If π and σ are related by τ_2, then σ is a snow leopard permutation.*

Proof. It turns out that (i) and (ii) depend on each other, so we prove them together.

It's routine to check that (i) and (ii) hold when π and σ have lengths -1, 1, or 3, so suppose $|\pi| = |\sigma| \geq 5$; we argue by induction on $|\pi|$.

Case One: π and σ are related by τ_1. By Theorem 2.20, there are snow leopard permutations π_1 and π_2 such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$.

First suppose $\pi_1 = @$, so that $\pi = 1 \ominus 1 \ominus \pi_2$. In this case, if $i \geq 3$ then our swap takes place inside π_2, so there is a permutation σ_2 which is related to π_2 by τ_1 such that $\sigma = 1 \ominus 1 \ominus \sigma_2$. By induction, σ_2 is a snow leopard permutation, so σ is also a snow leopard permutation by Theorem 2.20. On the other hand, if $i \leq 2$ then $i = 1$, since the first entry of π is odd and the second is even. In this case there is a permutation β of odd length such that $\pi = 1 \ominus 1 \ominus \cdots \ominus 1 \ominus \beta$, and β is a snow leopard permutation by Lemma 4.4. Now $\sigma = (1 \oplus \alpha^c \oplus 1) \ominus 1 \ominus \beta$, where α is an identity permutation of odd length. Since α and β are snow leopard permutations, σ is also a snow leopard permutation by Theorem 2.20.

Now suppose $\pi_1 \neq @$. In this case our decreasing sequence must be entirely contained in either π_1^c or $1 \ominus \pi_2$. Since the $1 \ominus \pi_2$ part of π begins with an even number, any decreasing sequence beginning with an odd number in this part of π must be contained in π_2. Therefore there is a permutation σ_2 which is related to π_2 by τ_1 such that $\sigma = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \sigma_2$. By induction, σ_2 is a snow leopard permutation, so σ is a snow leopard permutation by Theorem 2.20.

On the other hand, if our decreasing sequence is contained in π_1^c, then it corresponds to an increasing sequence in π_1 which begins with an even number. Therefore, there is a permutation σ_1 which is related to π_1 by τ_2, for which $\sigma = (1 \oplus \sigma_1^c \oplus 1) \ominus 1 \ominus \pi_2$. By induction, σ_1 is a snow leopard permutation, so σ is also a snow leopard permutation by Theorem 2.20.

Case Two: π and σ are related by τ_2. By Theorem 2.20, there are snow leopard permutations π_1 and π_2 such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. In addition, any increasing sequence in π must be entirely contained in the $1 \oplus \pi_1^c \oplus 1$ part of π, or in the π_2 part of π. If our increasing sequence is contained in the π_2 part of π, then there is a permutation σ_2 which is related to π_2 by τ_2, such that $\sigma = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \sigma_2$. By induction, σ_2 is a snow leopard permutation, so σ is also a snow leopard permutation by Theorem 2.20.

On the other hand, if our increasing sequence is contained in the $1 \oplus \pi_1^c \oplus 1$ part of π, then we must have $i \geq 2$ and $i \leq |\pi_1| + 1$, since this part of π begins and ends with odd numbers. That is, our increasing sequence must be entirely contained in π_1^c. Therefore, this increasing sequence corresponds to a decreasing sequence in π_1, all of whose entries have opposite parity with the corresponding entries in π. This means there is a permutation σ_1 which is related to π_1 by τ_1 such that $\sigma = (1 \oplus \sigma_1^c \oplus 1) \ominus 1 \ominus \pi_2$. By induction, σ_1 is a snow leopard permutation, so σ is also a snow leopard permutation by Theorem 2.20. □

We are interested in permutations which are connected by chains of permutations in which consecutive permutations are related by τ_1 or τ_2, so we make the following definition.

Definition 4.6. We say permutations π and σ are *τ-related* whenever there is a sequence $\alpha_1, \ldots, \alpha_n$ of permutations such that $\pi = \alpha_1$, $\sigma = \alpha_n$, and for each j, the permutations α_j and α_{j-1} are related by τ_1 or related by τ_2.

We can now show that the snow leopard permutations of odd length are exactly those permutations that are τ-related to the reverse identity.

Theorem 4.7. *A permutation π of length $2n-1$ is a snow leopard permutation if and only if it is τ-related to the decreasing permutation of length $2n-1$.*

Proof. ($\Rightarrow$) It is routine to verify this result when π has length -1, 1, or 3, so suppose $|\pi| \geq 5$; we argue by induction on $|\pi|$. By Theorem 2.20, there are snow

length	1	3	5	7	9
SLP-like permutations	1	2	7	32	175
SLPs	1	2	5	14	42

Table 3. The number of SLPs compared with the number of permutations with some properties of SLPs.

leopard permutations π_1 and π_2 of odd length such that $\pi = (1 \oplus \pi_1^c \oplus 1) \ominus 1 \ominus \pi_2$. By induction, there is a sequence s_1 (resp. s_2) of moves of types τ_1 and τ_2 which, when applied to the decreasing permutation of the appropriate length, produces π_1 (resp. π_2). To obtain π from the decreasing permutation of length $2n-1$, first apply a move of type τ_1 to swap the entries in positions 1 and $|\pi_1|+2$. Now apply the sequence s_2 of moves to the entries to the right of position $|\pi_1|+3$. Finally, for each move in s_1 of type τ_1, apply the corresponding move of type τ_2 to the subsequence in positions 2 through $|\pi_1|+1$, and vice versa. Since we have constructed each of the pieces of π individually, the resulting permutation is π itself.

($\Leftarrow$) It is routine to check that the decreasing permutation of length $2n-1$ is a snow leopard permutation, so this part is immediate from Theorem 4.5. □

Corollary 4.8. *Suppose π and σ are τ-related permutations of odd length. Then π is a snow leopard permutation if and only if σ is a snow leopard permutation.*

Proof. This is immediate from Theorem 4.7, since π and σ are snow leopard permutations if and only if they are τ-related to the decreasing permutation of length $|\pi|$, and this relationship is transitive. □

5. Questions and open problems

It should be possible to build on this work in a variety of directions. For example, it may be fruitful to study the distributions of various permutation statistics on snow leopard permutations, and to look for connections between these statistics and statistics on Catalan paths, or on other Catalan objects. In addition, both κ and the compatibility relation deserve more attention. Finally, we have the following more specific questions.

(1) Can we characterize the snow leopard permutations nonrecursively?

We have given a recursive decomposition of the snow leopard permutations, so in principle we can recognize these permutations in the wild using this decomposition. Similarly, we have also characterized the snow leopard permutations as the permutations generated by a particular set of transpositions. While these points of view are useful, we would also like to have a short list of simple conditions we can check to determine whether a given permutation is an SLP. For example, we know that if π is a snow leopard permutation of odd length then π preserves parity, π

avoids 2–14–3 and 3–41–2, and $\kappa(\pi)$ is a Catalan path. These conditions rule out many permutations, but there are still permutations with all of these properties which are not SLPs. In fact, in Table 3 we see how the number of permutations with these three properties compares with the number of snow leopard permutations for small lengths.

(2) What permutations of length n are compatible with alternating Baxter permutations of length $n+1$?

Cori, Dulucq, and Viennot [Cori et al. 1986] have used bijections with binary trees to prove that the alternating Baxter permutations of lengths $2n$ and $2n+1$ are counted by the products C_n^2 and C_nC_{n+1} of Catalan numbers, respectively. We conjecture that the smaller permutations which are compatible with the alternating Baxter permutations are counted by the same products of Catalan numbers. Our preliminary explorations suggest that we can extend either the work of Cori, Dulucq, and Viennot or the work of Dulucq and Guibert [1998] to prove this conjecture, but it might also be possible to extend or modify κ to give a proof.

Acknowledgements

We would like to thank Jim Propp for suggesting that we look into Canary's work on compatible ASMs and domino tilings of Aztec diamonds, and also for meeting with us to discuss this research during his visit to Carleton in November of 2013. We would also like to thank all members of the Carleton Mathematics and Statistics department for their support and wisdom.

References

[Ackerman et al. 2006] E. Ackerman, G. Barequet, and R. Y. Pinter, "A bijection between permutations and floorplans, and its applications", *Discrete Appl. Math.* **154**:12 (2006), 1674–1684. MR 2007a:05001 Zbl 1096.05002

[Asinowski 2014] A. Asinowski, "On permutation tilings of Aztec diamonds", personal communication, 2014.

[Asinowski et al. 2013] A. Asinowski, G. Barequet, M. Bousquet-Mélou, T. Mansour, and R. Y. Pinter, "Orders induced by segments in floorplans and (2-14-3, 3-41-2)-avoiding permutations", *Electron. J. Combin.* **20**:2 (2013), Paper #P35. MR 3084577 Zbl 1267.05018

[Baxter 1964] G. Baxter, "On fixed points of the composite of commuting functions", *Proc. Amer. Math. Soc.* **15** (1964), 851–855. MR 32 #1690 Zbl 0126.38701

[Boyce 1967] W. M. Boyce, "Generation of a class of permutations associated with commuting functions", *Math. Algorithms* **2** (1967), 19–26. Addendum op. cit., **3** (1968), 25–26. MR 40 #3750 Zbl 0266.05009

[Bressoud 1999] D. M. Bressoud, *Proofs and confirmations: the story of the alternating sign matrix conjecture*, Mathematical Association of America/Cambridge University Press, Washington, DC/Cambridge, 1999. MR 2000i:15002 Zbl 0944.05001

[Canary 2010] H. Canary, "Aztec diamonds and Baxter permutations", *Electron. J. Combin.* **17**:1 (2010), Paper #R105. MR 2012a:05023 Zbl 1201.52019

[Chung et al. 1978] F. R. K. Chung, R. L. Graham, V. E. Hoggatt, Jr., and M. Kleiman, "The number of Baxter permutations", *J. Combin. Theory Ser. A* **24**:3 (1978), 382–394. MR 82b:05011 Zbl 0398.05003

[Cori et al. 1986] R. Cori, S. Dulucq, and G. Viennot, "Shuffle of parenthesis systems and Baxter permutations", *J. Combin. Theory Ser. A* **43**:1 (1986), 1–22. MR 87k:05010 Zbl 0662.05004

[Dulucq and Guibert 1996] S. Dulucq and O. Guibert, "Stack words, standard tableaux and Baxter permutations", *Discrete Math.* **157**:1-3 (1996), 91–106. MR 98a:05009 Zbl 0870.05077

[Dulucq and Guibert 1998] S. Dulucq and O. Guibert, "Baxter permutations", *Discrete Math.* **180**:1-3 (1998), 143–156. MR 99c:05004 Zbl 0895.05002

[Elkies et al. 1992] N. Elkies, G. Kuperberg, M. Larsen, and J. Propp, "Alternating-sign matrices and domino tilings, I", *J. Algebraic Combin.* **1**:2 (1992), 111–132. MR 94f:52035 Zbl 0779.05009

[Guibert and Linusson 2000] O. Guibert and S. Linusson, "Doubly alternating Baxter permutations are Catalan", *Discrete Math.* **217**:1-3 (2000), 157–166. MR 2001i:05009 Zbl 0965.05005

[Mallows 1979] C. L. Mallows, "Baxter permutations rise again", *J. Combin. Theory Ser. A* **27**:3 (1979), 394–396. MR 81j:05014 Zbl 0427.05005

[Ouchterlony 2006] E. Ouchterlony, "Pattern avoiding doubly alternating permutations", pp. 652–663 in *Proceedings of the 18th Annual International Conference on Formal Power Series and Algebraic Combinatorics* (San Diego, CA, 2006), edited by M. Mishna, 2006.

[Propp 2001] J. Propp, "The many faces of alternating-sign matrices", pp. 43–58 in *Discrete models: combinatorics, computation, and geometry* (Paris, 2001), edited by R. Cori et al., Maison Inform. Math. Discrèt., Paris, 2001. MR 2003a:05038 Zbl 0990.05020

[Robbins 1991] D. P. Robbins, "The story of $1, 2, 7, 42, 429, 7436, \cdots$", *Math. Intelligencer* **13**:2 (1991), 12–19. MR 92d:05010 Zbl 0723.05004

[Stanley 1999] R. P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, 1999. MR 2000k:05026 Zbl 0928.05001

[Stanley 2013] R. P. Stanley, "Catalan addendum", electronic resource, May 25, 2013, http://www-math.mit.edu/~rstan/ec/catadd.pdf.

Received: 2014-04-28 Revised: 2014-12-05 Accepted: 2014-12-08

bjc.caffrey@gmail.com
Epic, 1979 Milky Way, Verona, WI 53593, United States

eegge@carleton.edu
Department of Mathematics and Statistics, Carleton College, Northfield, MN 55057, United States

miche417@umn.edu
Department of Mathematics, University of Minnesota Twin Cities, Minneapolis, MN 55455, United States

kailee.rubin@gmail.com
Epic, 1979 Milky Way Verona, WI 53593, United States

jonathanversteegh@gmail.com
Carleton College, Northfield, MN 55057, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.859

The Weibull distribution and Benford's law

Victoria Cuff, Allison Lewis and Steven J. Miller

(Communicated by John C. Wierman)

Benford's law states that many data sets have a bias towards lower leading digits (about 30% are 1s). It has numerous applications, from designing efficient computers to detecting tax, voter and image fraud. It's important to know which common probability distributions are almost Benford. We show that the Weibull distribution, for many values of its parameters, is close to Benford's law, quantifying the deviations. As the Weibull distribution arises in many problems, especially survival analysis, our results provide additional arguments for the prevalence of Benford behavior. The proof is by Poisson summation, a powerful technique to attack such problems.

1. Introduction to and applications of Benford's law

For any positive number x and base B, we can represent x in scientific notation as $x = S_B(x) \cdot B^{k(x)}$, where $S_B(x) \in [1, B)$ is called the significand[1] of x and the integer $k(x)$ represents the exponent. Benford's law of leading digits proposes a distribution for the significands which holds for many data sets, and states that the proportion of values beginning with digit d is approximately

$$\text{Prob}(\text{first digit is } d \text{ base } B) = \log_B\left(\frac{d+1}{d}\right). \tag{1-1}$$

More generally, the proportion with significand at most s base B is

$$\text{Prob}(1 \le S_B \le s) = \log_B s. \tag{1-2}$$

MSC2010: primary 60F05, 11K06; secondary 60E10, 42A16, 62E15, 62P99.

Keywords: Benford's law, Weibull distribution, digit bias, Poisson summation.

Cuff and Lewis were supported by NSF Grant DMS0850577 and Williams College; Miller was supported by NSF Grant DMS0970067 and DMS1265673. This work was done at the 2010 SMALL REU at Williams College, and a summary of it will appear in Chapter Three of *The Theory and Applications of Benford's Law*, to be published by Princeton University Press and edited by Miller. Since this work was written many of these results have been independently derived and applied to Internet traffic; see the work of Arshadi and Jahangir [2014].

[1] The significand is sometimes called the mantissa; however, such usage is discouraged by the IEEE and others, as mantissa is used for the fractional part of the logarithm, a quantity which is also important in studying Benford's law.

In particular, in base 10 the probability that the first digit is a 1 is about 30.1% (and not the 11% one would expect if each digit from 1 to 9 were equally likely).

This leading digit irregularity was first discovered by Newcomb [1881], who noticed that the earlier pages in the logarithmic books were more worn than other pages. Fifty years later, Benford [1938] observed the same digit bias in a variety of data sets. Benford studied the distribution of the first digits of 20 sets of data with over 20,000 total observations, including river lengths, populations, and mathematical sequences. For a full history and description of the law, see [Hill 1998; Raimi 1976], or go to [Berger and Hill 2012] or [Miller 2015] for additional reading.

One of the most fascinating aspects of Benford's law is the large and diverse list of fields studying it (auditing, computer science, dynamical systems, engineering, number theory, and statistics, to list a few). There are numerous applications, especially in fraud and data integrity. Two of the more famous are detecting tax and voter fraud [Cho and Gaines 2007; Mebane 2006; Nigrini 1996; 1997], but there are also applications in many other fields, ranging from round-off errors in computer science [Knuth 1997] to detecting image fraud and compression in engineering [Abdallah et al. 2015]. Already Benford's law has led to a variety of tests, either to detect fraud (in everything from corporate returns to medical studies) or to test data integrity; see, for example, [Judge and Schechter 2009; Nigrini 1997; Miller and Nigrini 2009].

In the next section we discuss attempts to explain the prevalence of Benford's law; unfortunately, some of these approaches are flawed, and have been incorrectly used for decades. Our purpose in this article is to highlight techniques from Fourier analysis that may not be widely known to the diverse group of researchers and aficionados in the field, emphasizing how Poisson summation provides a clean and correct way to quantify deviations from Benford's law for a variety of phenomena. Our main result is to quantify how close Weibull distributions are to Benford (we state these in Theorem 4.1 in Section 4, after first reviewing the needed prerequisites in Section 3; the proof is given in Section 5). For certain values of the scale and shape parameter, these distributions are almost Benford; this is quite important, as many survival distributions are modeled by Weibull distributions, and thus Benford tests are applicable.

2. Explanations of Benford's law

There have been numerous attempts to pass from observing the prevalence of Benford's law to explaining its occurrence in different and diverse systems. Such knowledge gives us a deeper understanding of which natural data sets should follow Benford's law. One of the earliest and most popular is due to Feller [1966], and has been the subject of many articles and papers since (a very good, recent description of this approach is given in [Fewster 2009]). It suggests that Benford behavior arises when a probability distribution is spread out over several orders of magnitude. Unfortunately, while some distributions satisfying this condition are close to Benford, others are not, and the method is sadly fundamentally flawed. See [Berger and Hill 2010; 2011b; Hill 2011] for detailed critiques of this method. The first rigorous explanation of Benford's law is due to Hill [1995] through scale invariance and measure theory (essentially, the distribution of leading digits should be invariant if we change scale); see also [Berger and Hill 2011a].

Rather than trying to prove why so many different phenomena are almost Benford, another approach is to study specific, important instances. In particular, there is an extensive literature on the leading digits of random variables and products of random variables of specific distributions (see for example [Miller and Nigrini 2008a]). While these arguments cannot be as general, the systems described arise in many important applications, making the importance of these researches clear.

The starting point of this work is the paper by Leemis, Schmeiser, and Evans [Leemis et al. 2000], who champion this viewpoint. They ran numerical simulations on a variety of parametric survival distributions to examine conformity to Benford's law. Among these distributions was the Weibull distribution, whose density is

$$f(x;\alpha,\gamma)=\begin{cases}(\gamma/\alpha)(x/\alpha)^{(\gamma-1)}\exp(-(x/\alpha)^{\gamma}) & \text{if } x\geq 0,\\ 0 & \text{otherwise,}\end{cases} \tag{2-1}$$

where $\alpha,\gamma>0$. Note that α adjusts the scale of the data and only γ affects the shape of the distribution.[2] Special cases of the Weibull distribution include the exponential distribution ($\gamma=1$) and the Rayleigh distribution ($\gamma=2$). The most common use of the Weibull distribution is in survival analysis, where a random variable X modeled by the Weibull distribution represents the "time-to-failure", resulting in a distribution where the failure rate is modeled relative to a power of time.

The Weibull distribution arises in problems in such diverse fields as food contents, engineering, medical data, politics, pollution and sabermetrics, along with many others; see [Carroll 2003; Corzo and Bracho 2008; Fry 2004; McShane et al. 2008; Mikolaj 1972; Miller 2007; Terawaki et al. 2006; Weibull 1951; Yiannoutsos 2009; Zhao et al. 2011] to name just a few. As the extensiveness of this list indicates, many data sets follow a Weibull distribution, and thus if we are going test for fraud or data integrity, it is essential to quantify how close these distributions are to Benford. Our goal in this work is to provide proofs of the observations of Leemis, Schmeiser, and Evans [Leemis et al. 2000] that Weibull distributions are often close to Benford, emphasizing the ideas behind the method, as these are applicable to a variety of other problems (see, for example, [Jang et al. 2009; Kontorovich and Miller 2005; Miller and Nigrini 2008b]).

[2] One could introduce another parameter, β, which would represent a translation of the data. Doing so replaces x with $x-\beta$, and the condition $x\geq 0$ becomes $x\geq\beta$. In this paper we concentrate on the case $\beta=0$.

3. Mathematical preliminaries

Our analysis generalizes the work of [Miller and Nigrini 2008b], where the exponential case was studied in detail (see also [Dümbgen and Leuenberger 2008] for another approach to analyzing exponential random variables). The main ingredients come from Fourier analysis, in particular, applying Poisson summation to the derivative of the cumulative distribution function of the logarithms modulo 1, F_B. We first review some needed definitions, then describe why it is so useful to study the logarithms modulo 1, and conclude with a quick review of Poisson summation.

(1) The gamma function $\Gamma(s)$ generalizes the factorial function; for n a nonnegative integer, we have $\Gamma(n+1) = n!$, and for $\Re(s) > 0$, we have

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1}\, dx$$

(we will need to evaluate the gamma function at complex arguments in our analysis); here $\Re(z)$ denotes the real part of z. See [Whittaker and Watson 1996] for an introduction and proofs of needed properties.

(2) We say a is congruent to b modulo 1 if $a-b$ is an integer; we denote this by $a = b \bmod 1$.

(3) A sequence $\{a_n\}_{n=1}^\infty \subset [0,1]$ is equidistributed if

$$\lim_{N\to\infty} \frac{\#\{n : n \le N, a_n \in [a,b]\}}{N} = b-a$$

for all $[a,b] \subset [0,1]$. Similarly a continuous random variable on $[0,\infty)$ whose probability density function is p is equidistributed modulo 1 if

$$\lim_{T\to\infty} \frac{\int_0^T \chi_{a,b}(x) p(x)\, dx}{\int_0^T p(x)\, dx} = b-a$$

for any $[a,b] \subset [0,1]$, where $\chi_{a,b}(x) = 1$ for $x \bmod 1 \in [a,b]$ and 0 otherwise.

(4) If f is an integrable function (so $\int_{-\infty}^{\infty} |f(x)|\, dx < \infty$) then its Fourier transform, denoted $\hat{f}$, is given by

$$\hat{f}(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x y}\, dx, \quad \text{where } e^{iu} = \cos u + i \sin u.$$

Note if X is a random variable with density f then this is a rescaled version of its characteristic function, $\mathbb{E}[e^{itX}]$.

(5) Let $\eta > 0$. We say f decays like $x^{-(1+\eta)}$ if there are constants $x_0, C_\eta > 0$ such that $|f(x)| \le C_\eta |x|^{-(1+\eta)}$ for all $|x| > x_0$.

One of the most common ways to prove a system is Benford is to show that its logarithms modulo 1 are equidistributed. We quickly sketch the proof of this equivalence; see [Diaconis 1977; Miller and Nigrini 2008b; Miller and Takloo-Bighash 2006] for details. If $y_n = \log_B x_n \bmod 1$ (thus y_n is the fractional part of the logarithm of x_n), then the significands of B^{y_n} and $x_n = B^{\log_B x_n}$ are equal, as these two numbers differ by a factor of B^k for some integer k. If now $\{y_n\}$ is equidistributed modulo 1, then by definition for any $[a,b] \subset [0,1]$, we have $\lim_{N\to\infty} \#\{n \le N : y_n \in [a,b]\}/N = b-a$. Taking $[a,b] = [0, \log_B s]$ implies that as $N \to \infty$, the probability that $y_n \in [0, \log_B s]$ tends to $\log_B s$, which by exponentiating implies that the probability that the significand of x_n is in $[1,s]$ tends to $\log_B s$, the Benford probability.

Given a random variable X, let F_B denote the cumulative distribution function of $\log_B X \bmod 1$. The above discussion shows that Benford's law is equivalent to $F_B(z) = z$, or our original random variable X is Benford if $F_B'(z) = 1$. This suggests that a natural way to investigate deviations from Benford behavior is to compare the deviation of $F_B'(z)$ from 1, which would represent a uniform distribution.

Fourier analysis is ideally suited for these computations. The reason is that in general one cannot throw away part of a mathematical expression and still maintain equality. For example, note $\sqrt{(x \bmod 1) + (y \bmod 1)}$ is neither equal to nor congruent modulo 1 to $\sqrt{x+y}$; however, $e^{2\pi i x}$ does equal $e^{2\pi i (x \bmod 1)}$. By using the complex exponentials, it is harmless to drop modulo 1 restrictions. As these restrictions naturally arise in investigating the first digit, it is natural to attack the problem with Fourier techniques.

The last ingredient we need is Poisson summation. We don't state it in its most general form, as the following weak version typically suffices for Benford investigations due to the smoothness of the underlying densities. See [Miller and Takloo-Bighash 2006] or [Stein and Shakarchi 2003] for a proof.

Theorem 3.1 (Poisson summation). *Let f, f' and f'' be continuous functions which decay like $x^{-(1+\eta)}$ for some $\eta > 0$. Then*

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n).$$

Our assumptions about f imply that $\hat{f}$ decays rapidly. The power of Poisson summation is that it typically allows us to exchange a slowly converging sum with a rapidly converging sum. In many applications only the $n = 0$ term matters; if f is a probability density then it integrates to 1, and hence $\hat{f}(0) = 1$. For us, this is important as it implies a sum over nonzero n can measure a deviation.

For example, consider the density of a normal random variable Y with mean 0 and variance $N/2\pi$; this example is very important in showing Brownian motions and many products of independent random variables become Benford (see [Miller and Takloo-Bighash 2006; Miller and Nigrini 2008a]). If we want to see how often $Y \bmod 1$ is in an interval $[a, b] \subset [0, 1]$, we need to study $\text{Prob}(Y \bmod 1 \in [a, b]) = \sum_{n=-\infty}^{\infty} \text{Prob}(Y \in [a+n, b+n])$. We *sketch* how Poisson summation enters, and provide full details when we prove our main result. The latter probabilities are integrals of the density over the intervals $[a+n, b+n]$, and if N is large each of these is approximately $b-a$ times the density at n. By Poisson summation, summing the density over n is the same as summing the Fourier transform at n:

$$\sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{N}} e^{-\pi n^2/N} = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 N}.$$

Note the sharp contrast between the two sums. For the first sum, all n with $|n| \le \sqrt{N}$ contribute the same order of magnitude, while for the second sum, the $n = 0$ term contributes 1 and the next term is immensely smaller (by a factor of $e^{-\pi N}$). This example illustrates how Poisson summation allows us to replace a slowly decaying sum of a density with a rapidly decaying one.

4. Main results

Our main result is the following extension of results for the exponential distribution, which measures the deviation of the logarithm modulo 1 of the Weibull distribution and the uniform distribution. It's thus not surprising that for γ close to 1, the digits are close to Benford, as $\gamma = 1$ corresponds to the exponential distribution. The main contribution below is quantifying how the fit worsens as γ grows. The larger γ is, the worse the fit. The effect of α is easier to explain. As the result of replacing α by αB is simply to rescale our random variable by a factor of B, the significand is unaffected. Thus it suffices to study α in the window $[1, B)$, but γ may be any real value.

Theorem 4.1. *Let $Z_{\alpha,\gamma}$ be a random variable whose density is Weibull with parameters $\alpha, \gamma > 0$ arbitrary. For $z \in [0, 1]$, let $F_B(z)$ be the cumulative distribution function of $\log_B Z_{\alpha,\gamma} \bmod 1$; thus $F_B(z) := \text{Prob}(\log_B Z_{\alpha,\gamma} \bmod 1 \in [0, z])$. Then the density of $\log_B Z_{\alpha,\gamma} \bmod 1$, $F'_B(z)$, is given by*

$$F'_B(z) = 1 + 2\sum_{m=1}^{\infty} \Re\left(e^{-2\pi i m(z - \log\alpha/\log B)}\,\Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right)\right). \tag{4-1}$$

In particular, the densities of $\log_B Z_{\alpha,\gamma} \bmod 1$ and $\log_B Z_{\alpha B,\gamma} \bmod 1$ are equal, and thus it suffices to consider only α in an interval of the form $[a, aB)$ for any $a > 0$.

From the fundamental equivalence, a straightforward integration immediately translates (4-1) into quantifying differences in the distribution of leading digits of Weibull random variables and Benford's law. Specifically, the probability of a first digit of d is obtained by integrating $F_B'(z)$ from $\log_B d$ to $\log_B(d+1)$. The main term comes from the constant 1, and is $\log_B((d+1)/d)$, the Benford probability; we discuss the size of the error in Theorem 4.2.

The above theorem is proved in the next section. As in [Miller and Nigrini 2008b], the proof involves applying Poisson summation to the derivative of the cumulative distribution function of the logarithms modulo 1, which as discussed in the previous section is a natural way to compare deviations from the resulting distribution and the uniform distribution. The key idea is that if a data set satisfies Benford's law, then the distribution of its logarithms will be uniform. Our series expansions are obtained by applying properties of the gamma function.

As the deviations of $F_B'(z)$ from being identically 1 measure the deviations from Benford behavior, it is important to have good estimates for the sum over m in (4-1). The bounds below have not been optimized, but instead have been chosen to simplify the algebra in the proofs (given in the Appendix). Thus we assume k below is at least 6, which is essentially equivalent to only investigating the case where the error ϵ is required to be of at most modest size (which is reasonable, as a series expansion with a large error is useless).

Theorem 4.2. *Let $F_B'(z)$ be as in (4-1).*

(1) *For $M \geq (\gamma \log B \log 2)/4\pi^2$, the error from dropping the $m \geq M$ terms in $F_B'(z)$ is at most*

$$\frac{2\sqrt{2}(\pi^2+\gamma\log B)\sqrt{\gamma\log B}}{\pi^3} M e^{-\pi^2 M/(\gamma \log B)}.$$

(2) *In order to have an error of at most ϵ in evaluating $F_B'(z)$, it suffices to take the first M terms, where $M = (k + \ln k + 1/2)/a$, with $k = \max(6, -\ln(a\epsilon/C))$, $a = \pi^2/(\gamma \log B)$, and*

$$C = \frac{2\sqrt{2}(\pi^2+\gamma\log B)\sqrt{\gamma\log B}}{\pi^3}.$$

For further analysis, we compared our series expansion for the derivative to the uniform distribution through a Kolmogorov–Smirnov test; see Figure 1 for a contour plot of the discrepancy. This statistic measures the absolute value of the greatest difference in cumulative distribution functions of two densities. Thus the larger the value, the further apart they are. Note the good fit observed between the two distributions when $\gamma = 1$ (representing the exponential distribution), which has already been proven to be a close fit to the Benford distribution ([Dümbgen and Leuenberger 2008; Leemis et al. 2000; Miller and Nigrini 2008b]).

The Kolmogorov–Smirnov metric gives a good comparison because it allows us to compare the distributions in terms of both parameters, γ and α. We also look at two other measures of closeness, the L_1-norm and the L_2-norm, both of which

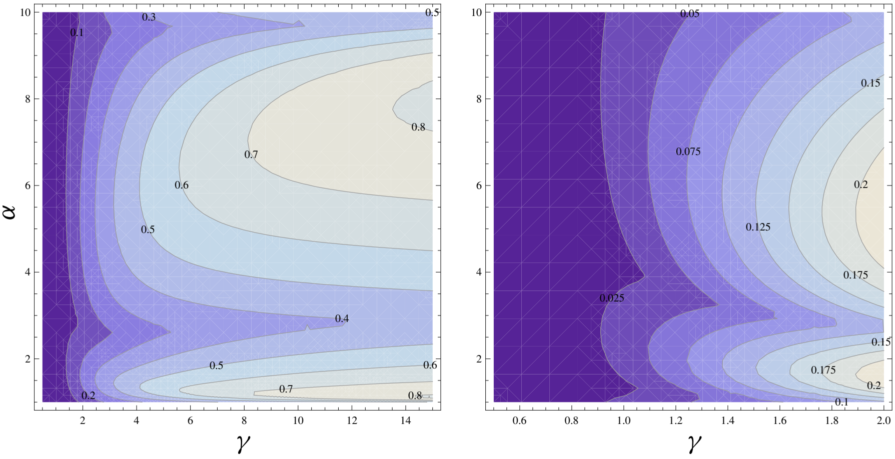


Figure 1. Kolmogorov–Smirnov test. Left: $\gamma \in [0, 15]$. Right: $\gamma \in [.5, 2]$. As γ (the shape parameter on the horizontal axis) increases, the Weibull distribution is no longer a good fit compared to the uniform. Note that α (the scale parameter on the vertical axis) has less of an effect on the overall conformance.

also test the differences between (4-1) and the uniform distribution; see Figure 2. The L_1-norm of $f - g$ is $\int_0^1 |f(t) - g(t)|\,dt$, which puts equal weights on the all deviations, while the L_2-norm is given by $\int_0^1 |f(t) - g(t)|^2\,dt$, which unlike the L_1-norm puts more weight on larger differences. The closer γ is to zero the better the fit. As γ increases, the cumulative Weibull distribution is no longer a good fit compared to 1. The L_1- and L_2-norms are independent of α.

The combination of the Kolmogorov–Smirnov tests and the L_1- and L_2-norms show us that the Weibull distribution almost exhibits Benford behavior when γ is modest; as γ increases, the Weibull distribution no longer conforms to the expected leading digit probabilities. The scale parameter α does have a small

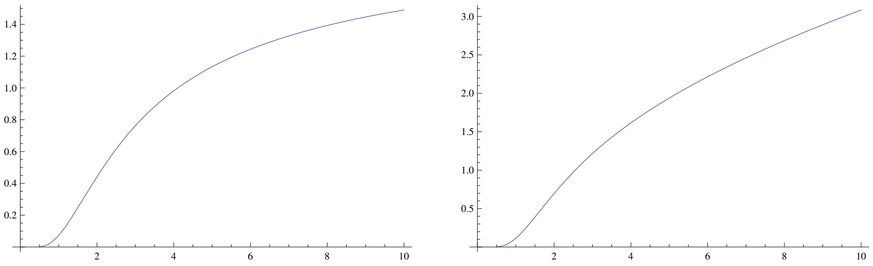


Figure 2. Left: L_1-norm of $F'_B(z) - 1$ for $\gamma \in [0.5, 10]$. Right: L_2-norm of $F'_B(z) - 1$ for $\gamma \in [0.5, 10]$.

effect on the conformance as well, but not nearly to the same extreme as the shape parameter γ. Fortunately in many applications the scale parameter γ is not too large (it is frequently less than 2 in the Weibull distribution references cited earlier), and thus our work provides additional support for the prevalence of Benford behavior.

5. Proof of main result

To prove Theorem 4.1, we study the distribution of $\log_B Z_{\alpha,\gamma} \bmod 1$ when $Z_{\alpha,\gamma}$ has the Weibull distribution with parameters α and γ. The analysis is aided by the fact that the cumulative distribution function for a Weibull random variable has a nice closed form expression; for $Z_{\alpha,\gamma}$, the cumulative distribution function is $\mathcal{F}_{\alpha,\gamma}(x) = 1 - \exp(-(x/a)^\gamma)$. Let $[a,b] \subset [0,1]$. Then

$$\begin{aligned}\operatorname{Prob}\big(\log_B Z_{\alpha,\gamma} \bmod 1 \in [a,b]\big) &= \sum_{k=-\infty}^{\infty} \operatorname{Prob}\big(\log_B Z_{\alpha,\gamma} \bmod 1 \in [a+k, b+k]\big) \\ &= \sum_{k=-\infty}^{\infty} \operatorname{Prob}\big(Z_{\alpha,\gamma} \in [B^{a+k}, B^{b+k}]\big) \\ &= \sum_{k=-\infty}^{\infty} \exp\left(-\left(\frac{B^{a+k}}{\alpha}\right)^{\gamma}\right) - \exp\left(-\left(\frac{B^{b+k}}{\alpha}\right)^{\gamma}\right).\end{aligned} \tag{5-1}$$

Proof of Theorem 4.1. It suffices to investigate (5-1) in the special case when $a = 0$ and $b = z$, since for any other interval $[a,b]$, we may determine its probability by subtracting the probability of $[0,a]$ from $[0,b]$. Thus, we study the cumulative distribution function of $\log_B Z_{\alpha,\gamma} \bmod 1$ for $z \in [0,1]$, which we denote by $F_B(z)$:

$$\begin{aligned}F_B(z) &:= \operatorname{Prob}\big(\log_B Z_{\alpha,\gamma} \bmod 1 \in [0,z]\big) \\ &= \sum_{k=-\infty}^{\infty} \exp\left(-\left(\frac{B^{k}}{\alpha}\right)^{\gamma}\right) - \exp\left(-\left(\frac{B^{z+k}}{\alpha}\right)^{\gamma}\right).\end{aligned} \tag{5-2}$$

This series expansion is rapidly converging, and the closeness of $Z_{\alpha,\gamma}$ to the Benford distribution is equivalent to the rapidly converging series in (5-2) for $F_B(z)$ being close to z for all z.

A natural way to investigate the closeness of $F_B(z)$ to z is to compare $F'(z)$ to 1. As in [Miller and Nigrini 2008b], studying the derivative $F'_B(z)$ is an easier way to approach this problem because we obtain a simpler Fourier transform than the Fourier transform of $e^{-(B^k/\alpha)^\gamma} - e^{-(B^{z+k}/\alpha)^\gamma}$. We then can analyze the obtained Fourier transform by applying Poisson summation (Theorem 3.1).

We use the fact that the derivative of the infinite sum $F_B(z)$ is the sum of the derivatives of the individual summands. This is justified by the rapid decay of summands, yielding

$$\begin{aligned} F_B'(z) &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha}\exp\biggl(-\biggl(\frac{B^{z+k}}{\alpha}\biggr)^{\gamma}\biggr) B^{z+k}\biggl(\frac{B^{z+k}}{\alpha}\biggr)^{\gamma-1}\gamma\log B \\ &= \sum_{k=-\infty}^{\infty} \frac{1}{\alpha}\exp\biggl(-\biggl(\frac{\zeta B^{k}}{\alpha}\biggr)^{\gamma}\biggr) \zeta B^{k}\biggl(\frac{\zeta B^{k}}{\alpha}\biggr)^{\gamma-1}\gamma\log B, \qquad (5\text{-}3) \end{aligned}$$

where for $z \in [0, 1]$, we use the change of variables $\zeta = B^z$.

We introduce

$$H(t) = \frac{1}{\alpha}\exp\biggl(-\biggl(\frac{\zeta B^{t}}{\alpha}\biggr)^{\gamma}\biggr) \zeta B^{t}\biggl(\frac{\zeta B^{t}}{\alpha}\biggr)^{\gamma-1}\gamma\log B,$$

where $\zeta \geq 1$ as $\zeta = B^z$ with $z \geq 0$. Since $H(t)$ is decaying rapidly we may apply Poisson summation; thus

$$\sum_{k=-\infty}^{\infty} H(k) = \sum_{k=-\infty}^{\infty} \widehat{H}(k), \qquad (5\text{-}4)$$

where $\widehat{H}$ is the Fourier Transform of $H : \widehat{H}(u) = \int_{-\infty}^{\infty} H(t)e^{-2\pi i t u}\,dt$. Therefore

$$\begin{aligned} F_B'(z) &= \sum_{k=-\infty}^{\infty} H(k) = \sum_{k=-\infty}^{\infty} \widehat{H}(k) \\ &= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\alpha}\exp\biggl(-\biggl(\frac{\zeta B^{t}}{\alpha}\biggr)^{\gamma}\biggr) \zeta B^{t}\biggl(\frac{\zeta B^{t}}{\alpha}\biggr)^{\gamma-1}\gamma e^{-2\pi i t k}\log B\,dt. \end{aligned} \qquad (5\text{-}5)$$

We change variables again, setting $w = (\zeta B^t/\alpha)^{\gamma}$, which implies

$$t = \log_B\biggl(\frac{\alpha w^{1/\gamma}}{\zeta}\biggr) \quad \text{and} \quad dw = \frac{1}{\alpha}\biggl(\frac{\zeta B^{t}}{\alpha}\biggr)^{\gamma-1}\zeta B^t\gamma\log B\,dt, \qquad (5\text{-}6)$$

so that

$$\begin{aligned} F_B'(z) &= \sum_{k=-\infty}^{\infty} \int_0^{\infty} e^{-w}\exp\biggl(-2\pi i k\log_B\biggl(\frac{\alpha w^{1/\gamma}}{\zeta}\biggr)\biggr)\,dw \\ &= \sum_{k=-\infty}^{\infty} \biggl(\frac{\alpha}{\zeta}\biggr)^{-2\pi i k/\log B} \int_0^{\infty} e^{-w} w^{-2\pi i k/(\gamma\log B)}\,dw \\ &= \sum_{k=-\infty}^{\infty} \biggl(\frac{\alpha}{\zeta}\biggr)^{-2\pi i k/\log B} \Gamma\biggl(1-\frac{2\pi i k}{(\gamma\log B)}\biggr), \qquad (5\text{-}7) \end{aligned}$$

where we used the definition of the gamma function in the last line. As $\Gamma(1) = 1$, we have

$$F_B'(z) = 1 + \sum_{m=1}^{\infty} \left(\frac{\zeta}{\alpha}\right)^{2\pi i m/\log B} \Gamma\left(1 - \frac{2\pi i m}{\gamma \log B}\right) + \left(\frac{\zeta}{\alpha}\right)^{-2\pi i m/\log B} \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right). \tag{5-8}$$

As in [Miller and Nigrini 2008b], the above series expansion is rapidly convergent. As $\zeta = B^z$, we have

$$\left(\frac{\zeta}{\alpha}\right)^{2\pi i m/\log B} = \cos\left(2\pi m z - 2\pi m\left(\frac{\log \alpha}{\log B}\right)\right) + i \sin\left(2\pi m z - 2\pi m\left(\frac{\log \alpha}{\log B}\right)\right), \tag{5-9}$$

which gives a Fourier series expansion for $F_B'(z)$ with coefficients arising from special values of the gamma function.

Using properties of the gamma function, we are able to improve (5-8). If $y \in \mathbb{R}$ then $\Gamma(1 - iy) = \overline{\Gamma(1 + iy)}$ (where the bar denotes complex conjugation). Thus the m-th summand in (5-8) is the sum of a number and its complex conjugate, which is simply twice the real part. We use the following standard relationship (see, for example, [Abramowitz and Stegun 1964]):

$$\left|\Gamma(1 + ix)\right|^2 = \frac{\pi x}{\sinh(\pi x)} = \frac{2\pi x}{e^{\pi x} - e^{-\pi x}}. \tag{5-10}$$

Writing the summands in (5-8) as

$$2\Re\left(e^{-2\pi i m(z - \log \alpha/\log B)}\, \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right)\right),$$

(5-8) becomes

$$F_B'(z) = 1 + 2\sum_{m=1}^{\infty} \Re\left(e^{-2\pi i m(z - \log \alpha/\log B)}\, \Gamma\left(1 + \frac{2\pi i m}{\gamma \log B}\right)\right). \tag{5-11}$$

Finally, in the exponential argument above, there is no change in replacing α with αB, as this changes the argument by $2\pi i$. Thus it suffices to consider $\alpha \in [a, aB)$ for any $a > 0$. □

This proof demonstrates the power of using Poisson summation in Benford's law problems, as it allows us to convert a slowly convergent series expansion into a rapidly converging one, with the main term corresponding to Benford behavior and the other terms measuring the deviation.

Appendix: Proofs of bounding estimates

We first estimate the contribution to $F_B'(z)$ from the tail, say from the terms with $m \geq M$. We do not attempt to derive the sharpest bounds possible, but rather highlight the method in a general enough case to provide useful estimates.

Proof of Theorem 4.2(1). We must bound the truncation error

$$\mathcal{E}_B(z) := \Re \sum_{m=M}^{\infty} e^{-2\pi i m(z-\log\alpha/\log B)}\,\Gamma\left(1+\frac{2\pi i m}{\gamma\log B}\right), \tag{A-1}$$

where $\Gamma(1+iu) = \int_0^\infty e^{-x}x^{iu}\,dx = \int_0^\infty e^{-x}e^{iu\log x}\,dx$. Note that in our case, $u = 2\pi m/(\gamma\log B)$. As u increases there is more oscillation and therefore more cancellation, resulting in a smaller value for our integral. Since $|e^{i\theta}| = 1$, if we take absolute values inside the sum, we have $|e^{-2\pi i m(z-\log\alpha/\log B)}| = 1$, and thus we may ignore this term in computing an upper bound.

Using standard properties of the gamma function, we have

$$\left|\Gamma(1+ix)\right|^2 = \frac{\pi x}{\sinh(\pi x)} = \frac{2\pi x}{e^{\pi x}-e^{-\pi x}}, \quad \text{where } x = \frac{2\pi m}{\gamma\log B}. \tag{A-2}$$

This yields

$$|\mathcal{E}_B(z)| \le \sum_{m=M}^{\infty} 1\left(\frac{4\pi^2 m}{\gamma\log B}\,\frac{1}{e^{2\pi^2 m/(\gamma\log B)}-e^{-2\pi^2 m/(\gamma\log B)}}\right)^{1/2}. \tag{A-3}$$

Let $u = e^{2\pi^2 m/(\gamma\log B)}$. We overestimate our error term by removing the difference of the exponentials in the denominator. Simple algebra shows that for

$$\frac{1}{u-\frac{1}{u}} \le \frac{2}{u},$$

we need $u \ge \sqrt{2}$. For us this means $e^{2\pi^2 m/(\gamma\log B)} \ge \sqrt{2}$, allowing us to simplify the denominator if $m \ge (\gamma\log B\log 2)/4\pi^2$, which we may do as we assumed M exceeds this value and $m \ge M$. We substitute this bound into (A-2), and replace $\sqrt{m}$ with m to simplify the resulting integral:

$$|\mathcal{E}_B(z)| \le \sum_{m=M}^{\infty}\left(\frac{4\pi^2 m}{\gamma\log B}\right)^{1/2}\frac{\sqrt{2}}{e^{\pi^2 m/(\gamma\log B)}} \le \frac{2\sqrt{2}\pi}{\sqrt{\gamma\log B}}\int_M^\infty m e^{-\pi^2 m/(\gamma\log B)}\,dm. \tag{A-4}$$

Letting $a = \pi^2/(\gamma\log B)$, integrating by parts gives

$$|\mathcal{E}_B(z)| \le \frac{2\sqrt{2}\pi}{\sqrt{\gamma\log B}}\,\frac{1}{a^2}(aMe^{-aM}+e^{-aM}) \le \frac{2\sqrt{2}\pi}{\sqrt{\gamma\log B}}\,\frac{a+1}{a^2}Me^{-aM} \tag{A-5}$$

(since $M \ge 1$, $aM+1 \le (a+1)M$), which after some algebra simplifies to

$$|\mathcal{E}_B(z)| \le \frac{2\sqrt{2}(\pi^2+\gamma\log B)\sqrt{\gamma\log B}}{\pi^3}Me^{-\pi^2 M/(\gamma\log B)}, \tag{A-6}$$

which is the error listed in Theorem 4.2(1). □

Proof of Theorem 4.2(2). Given the estimation of the error term from above, we now ask the related question: given an $\epsilon > 0$, how large must M be so that the first M terms give $F'_B(z)$ accurately to within ϵ of the true value? Let

$$C = \frac{2\sqrt{2}(\pi^2 + \gamma \log B)\sqrt{\gamma \log B}}{\pi^3}$$

and $a = \pi^2/(\gamma \log B)$. We must choose M so that $CMe^{-aM} \le \epsilon$, or equivalently

$$\frac{C}{a} aMe^{-aM} \le \epsilon. \tag{A-7}$$

As this is a transcendental equation in M, we do not expect a nice closed form solution, but we can obtain a closed form expression for a bound on M; for any specific choices of C and a, we can easily numerically approximate M. We let $u = aM$, giving

$$ue^{-u} \le a\epsilon/C. \tag{A-8}$$

With a further change of variables, we let $k = -\ln(a\epsilon/C)$ and then expand u as $u = k + x$ (as the solution should be close to k). We find

$$ue^{-u} \le e^{-k} \quad \text{is equivalent to} \quad \frac{k+x}{e^x} \le 1. \tag{A-9}$$

We try $x = \ln k + \frac{1}{2}$ and see

$$\frac{k+x}{e^x} \le 1 \quad \text{is equivalent to} \quad \frac{k + \ln k + \frac{1}{2}}{ke^{1/2}} \le 1. \tag{A-10}$$

From here, we want to determine the value of k such that $\ln k \le \frac{1}{2}k$, as this ensures the needed inequality above holds. Exponentiating, we need $k^2 \le e^k$. As $e^k \ge k^3/3!$ for k positive, it suffices to choose k so that $k^2 \le k^3/6$, or $k \ge 6$; this holds for ϵ sufficiently small. For $k \ge 6$, we have

$$k + \ln k + \tfrac{1}{2} \le k + \tfrac{1}{2}k + \tfrac{1}{12}k = \tfrac{19}{12}k \approx 1.5833k, \tag{A-11}$$

but

$$ke^{1/2} \approx 1.64872k. \tag{A-12}$$

Therefore a correct cutoff value for M, in order to have an error of at most ϵ, is

$$M = \frac{k + \ln k + \frac{1}{2}}{a}, \tag{A-13}$$

where

$$k = \max\left(k, -\ln\frac{a\epsilon}{C}\right), \quad a = \frac{\pi^2}{\gamma \log B}, \quad C = \frac{2\sqrt{2}(\pi^2 + \gamma \log B)\sqrt{\gamma \log B}}{\pi^3}. \tag{A-14}$$

□

References

[Abdallah et al. 2015] C. T. Abdallah, G. L. Heileman, S. J. Miller, F. Pérez-González, and T. Quach, "Application of Benford's law to images", in *Theory and applications of Benford's law*, edited by S. J. Miller, Princeton University Press, 2015.

[Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503

[Arshadi and Jahangir 2014] L. Arshadi and A. H. Jahangir, "Benford's law behavior of internet traffic", *J. Netw. Comput. Appl.* **40** (2014), 194–205.

[Benford 1938] F. Benford, "The law of anomalous numbers", *Proc. Amer. Phil. Soc.* **78**:4 (1938), 551–572. Zbl 0018.26502

[Berger and Hill 2010] A. Berger and T. P. Hill, "Fundamental flaws in Feller's classical derivation of Benford's law", preprint, 2010. arXiv 1005.2598

[Berger and Hill 2011a] A. Berger and T. P. Hill, "A basic theory of Benford's law", *Probab. Surv.* **8** (2011), 1–126. MR 2012h:37015 Zbl 1245.60016

[Berger and Hill 2011b] A. Berger and T. P. Hill, "Benford's law strikes back: no simple explanation in sight for mathematical gem", *Math. Intelligencer* **33**:1 (2011), 85–91. MR 2012j:62006 Zbl 1221.60010

[Berger and Hill 2012] A. Berger and T. P. Hill, "Benford online bibliography", 2012, available at http://www.benfordonline.net.

[Carroll 2003] K. J. Carroll, "On the use and utility of the Weibull model in the analysis of survival data", *Control. Clin. Trials* **24**:6 (2003), 682–701.

[Cho and Gaines 2007] W. K. T. Cho and B. J. Gaines, "Breaking the (Benford) law: statistical fraud detection in campaign finance", *Amer. Stat.* **61**:3 (2007), 218–223. MR 2393725

[Corzo and Bracho 2008] O. Corzo and N. Bracho, "Application of Weibull distribution model to describe the vacuum pulse osmotic dehydration of sardine sheets", *LWT Food Sci. Technol.* **41**:6 (2008), 1108–1115.

[Diaconis 1977] P. Diaconis, "The distribution of leading digits and uniform distribution mod 1", *Ann. Probability* **5**:1 (1977), 72–81. MR 54 #10178 Zbl 0364.10025

[Dümbgen and Leuenberger 2008] L. Dümbgen and C. Leuenberger, "Explicit bounds for the approximation error in Benford's law", *Electron. Commun. Probab.* **13** (2008), 99–112. MR 2009b:60056 Zbl 1189.60044

[Feller 1966] W. Feller, *An introduction to probability theory and its applications*, vol. 2, 2nd ed., Wiley, New York, 1966. MR 35 #1048 Zbl 0138.10207

[Fewster 2009] R. M. Fewster, "A simple explanation of Benford's law", *Amer. Stat.* **63**:1 (2009), 26–32. MR 2655700

[Fry 2004] S. Fry, "How political rhetoric contributes to the stability of coercive rule: a Weibull model of post-abuse government survival", 2004. Paper presented at the annual meeting of the International Studies Association (Montreal, 2004).

[Hill 1995] T. P. Hill, "A statistical derivation of the significant-digit law", *Statist. Sci.* **10**:4 (1995), 354–363. MR 98a:60021 Zbl 0955.60509

[Hill 1998] T. P. Hill, "The first digit phenomenon", *Amer. Sci.* **86**:4 (1998), 358–363.

[Hill 2011] T. P. Hill, "Benford's law blunders", *Amer. Stat.* **65**:2 (2011), 141.

[Jang et al. 2009] D. Jang, J. U. Kang, A. Kruckman, J. Kudo, and S. J. Miller, "Chains of distributions, hierarchical Bayesian models and Benford's law", *J. Alg. Number Theory Adv. Appl.* **1**:1 (2009), 37–60. Zbl 1180.11026 arXiv 0805.4226

[Judge and Schechter 2009] G. Judge and L. Schechter, "Detecting problems in survey data using Benford's law", *J. Hum. Resour.* **44**:1 (2009), 1–24.

[Knuth 1997] D. E. Knuth, *The art of computer programming, 2: Seminumerical algorithms*, 3rd ed., Addison-Wesley, Reading, MA, 1997. MR 83i:68003 Zbl 0895.65001

[Kontorovich and Miller 2005] A. V. Kontorovich and S. J. Miller, "Benford's law, values of L-functions and the $3x + 1$ problem", *Acta Arith.* **120**:3 (2005), 269–297. MR 2007c:11085 Zbl 1139.11033

[Leemis et al. 2000] L. M. Leemis, B. W. Schmeiser, and D. L. Evans, "Survival distributions satisfying Benford's law", *Amer. Stat.* **54**:4 (2000), 236–241. MR 1803620

[McShane et al. 2008] B. McShane, M. Adrian, E. T. Bradlow, and P. S. Fader, "Count models based on Weibull interarrival times", *J. Bus. Econom. Statist.* **26**:3 (2008), 369–378. MR 2009h:60095

[Mebane 2006] W. R. Mebane, Jr., "Election forensics: the second-digit Benford's law test and recent American presidential elections", 2006, available at http://www.umich.edu/ wmebane/fraud06.pdf. Paper presented at the Election Fraud Conference (Salt Lake City, UT, 2006).

[Mikolaj 1972] P. G. Mikolaj, "Environmental applications of the Weibull distribution function: oil pollution", *Science* **176**:4038 (1972), 1019–1021.

[Miller 2007] S. J. Miller, "A derivation of the Pythagorean won-loss formula in baseball", *Chance* **20**:1 (2007), 40–48. MR 2361359

[Miller 2015] S. J. Miller (editor), *Benford's law: theory and applications*, Princeton University Press, 2015. Zbl 06446729

[Miller and Nigrini 2008a] S. J. Miller and M. J. Nigrini, "The modulo 1 central limit theorem and Benford's law for products", *Int. J. Algebra* **2**:3 (2008), 119–130. MR 2009e:60053 Zbl 1148.60008

[Miller and Nigrini 2008b] S. J. Miller and M. J. Nigrini, "Order statistics and Benford's law", *Int. J. Math. Math. Sci.* **2008** (2008), Article ID 382948. MR 2010c:62168 Zbl 05534756 arXiv math/0601344

[Miller and Nigrini 2009] S. J. Miller and M. J. Nigrini, "Data diagnostics using second order tests of Benford's Law", *Audit. J. Pract. Theory* **28**:2 (2009), 305–324.

[Miller and Takloo-Bighash 2006] S. J. Miller and R. Takloo-Bighash, *An invitation to modern number theory*, Princeton University Press, 2006. MR 2006k:11002 Zbl 1155.11001

[Newcomb 1881] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers", *Amer. J. Math.* **4**:1-4 (1881), 39–40. MR 1505286 JFM 13.0161.01

[Nigrini 1996] M. J. Nigrini, "Digital analysis and the reduction of auditor litigation risk", pp. 68–81 in *Auditing symposium XIII: proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems* (Lawrence, KS, 1996), edited by M. L. Ettredge, Division of Accounting and Information Systems, School of Business, University of Kansas, Lawrence, KS, 1996.

[Nigrini 1997] M. J. Nigrini, "The use of Benford's law as an aid in analytical procedures", *Audit. J. Pract. Theory* **16**:2 (1997), 52–67.

[Raimi 1976] R. A. Raimi, "The first digit problem", *Amer. Math. Monthly* **83**:7 (1976), 521–538. MR 53 #14593 Zbl 0349.60014

[Stein and Shakarchi 2003] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Princeton Lectures in Analysis **1**, Princeton University Press, 2003. MR 2004a:42001 Zbl 1026.42001

[Terawaki et al. 2006] Y. Terawaki, T. Katsumi, and V. Ducrocq, "Development of a survival model with piecewise Weibull baselines for the analysis of length of productive life of Holstein cows in Japan", *J. Dairy Sci.* **89**:10 (2006), 4058–4065.

[Weibull 1951] W. Weibull, "A statistical distribution function of wide applicability", *J. Appl. Mech.* **18** (1951), 293–297. Zbl 0042.37903

[Whittaker and Watson 1996] E. T. Whittaker and G. N. Watson, *A course of modern analysis: an introduction to the general theory of infinite processes and of analytic functions, with an account of the principal transcendental functions*, Reprint of the 4th ed., Cambridge University Press, 1996. MR 97k:01072 Zbl 0951.30002

[Yiannoutsos 2009] C. T. Yiannoutsos, "Modeling AIDS survival after initiation of antiretroviral treatment by Weibull models with changepoints", *J. Int. AIDS Soc.* **12**:9 (2009).

[Zhao et al. 2011] Y. Zhao, A. H. Lee, K. K. W. Yau, and G. J. McLachlan, "Assessing the adequacy of Weibull survival models: a simulated envelope approach", *J. Appl. Stat.* **38**:10 (2011), 2089–2097. MR 2843245

Received: 2014-07-31 Revised: 2014-10-19 Accepted: 2014-12-01

vcuff@g.clemson.edu
Department of Mathematics, Clemson University,
Clemson, SC 29634, United States

allewis2@ncsu.edu
Department of Mathematics, North Carolina State University,
Raleigh, NC 27695, United States

sjm1@williams.edu
Department of Mathematics and Statistics, Williams College,
Williamstown, MA 01267, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.875

Differentiation properties of the perimeter-to-area ratio for finitely many overlapped unit squares

Paul D. Humke, Cameron Marcott, Bjorn Mellem and Cole Stiegler

(Communicated by Frank Morgan)

In this paper we examine finite unions of unit squares in same plane and consider the ratio of perimeter to area of these unions. In 1998, T. Keleti published the conjecture that this ratio never exceeds 4. Here we study the continuity and differentiability of functions derived from the geometry of the union of those squares. Specifically we show that if there is a counterexample to Keleti's conjecture, there is also one where the associated ratio function is differentiable.

1. Introduction

The purpose of this paper is to introduce several functions associated with the *perimeter-to-area conjecture* (*PAC*) of Tamás Keleti [1998] and to investigate the smoothness properties of those functions.

Keleti's perimeter-to-area conjecture (PAC). *The perimeter-to-area ratio of the union of finitely many unit squares in a plane does not exceed* 4.

The problem of showing such a ratio is bounded first seems to have appeared as Problem 6 in the 1998 edition of the famous Miklós Schweitzer Competition in Hungary [Competition 1998]. Later that same year, Keleti published his perimeter-to-area conjecture that this bound is actually 4. To date, the best known bound is slightly less than 5.6. This bound was achieved by Keleti's student Zoltán Gyenes [2005] in his master's thesis. A special case of the theorem, where all of the squares are axis oriented, is known to be true; Gyenes also presents a proof of this case in the above work, and the authors present two additional proofs in [Humke et al. 2015]. The PAC is particularly intriguing as some of its obvious generalizations are false. Gyenes [2005] showed that the corresponding ratio for unions of congruent convex sets need not be bounded by the ratio for a single copy of the set.

MSC2010: 26B05.
Keywords: Keleti, perimeter-to-area ratio.

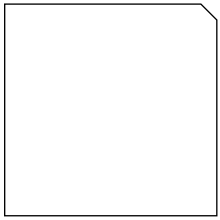


Figure 1. A convex set counterexample to the PAC for general convex sets.

Gyenes's example. There exist congruent convex sets $E_1 \cong E_2 \subset \mathbb{R}^2$ such that the perimeter-to-area ratio for $E_1 \cup E_2$ exceeds the perimeter-to-area ratio for either one of them.

The Gyenes example is disarmingly straightforward. The convex set template is an origin-centered unit square with one judiciously chosen isosceles corner triangle removed. That corner triangle is chosen so that the perimeter-to-area ratio of the resulting figure is less than 4. But the union of this template with a rotated copy is simply the original unit square whose perimeter-to-area ratio is exactly 4. See Figure 1.

In this paper, we build machinery for analyzing the PAC, showing that for almost all finite unions of squares, the perimeter to area ratio is differentiable in the usual Euclidean sense. If a counterexample exists, then there exists a counterexample where the derivative exists. These results provide inroads toward understanding the PAC by potentially relating it to large body of discrete geometric work, including the Kneser–Poulsen theorem and results by Ho-Lun Cheng and Herbert Edelsbrunner [2003] on derivatives when translating circles in the plane.

2. Notation and setting

Let $H = \bigcup_{i=1}^{n} H_i$ be the finite union of unit squares H_i in $\mathbb{R}^2$. Let the perimeter and area functions, $p(\,\cdot\,)$ and $\alpha(\,\cdot\,)$ respectively, take a closed, bounded polygonal figure in the plane as input and return that figure's perimeter and area respectively. If S is a set, we denote the boundary of S by $\operatorname{bd} S$. Throughout we will be interested in the boundary of polygonal regions, and one focus of our attention will be the (maximal) segments comprising that boundary. We refer to these maximal segments as *component segments* of the boundary. The ϵ-ball about a set S will be denoted by $B_\epsilon(S)$ and the convex hull of a set of points $\{p_i : i = 1, 2, \ldots, k\} \subset \mathbb{R}^2$ is denoted by $[p_1, p_2, \ldots, p_k]$.

Any point $(s_i, t_i, \phi_i)_{i=1}^{n} \in \mathbb{R}^{3n}$ may be mapped to an ordered union of n squares by taking (s_i, t_i) to be the rectangular coordinates of the center of the i-th square H_i and ϕ_i to be the smallest angle between the horizontal and a side of H_i. For notational convenience, we will also denote a single component square H_i by its coordinates, i.e., $H_i = (s_i, t_i, \phi_i)$. This correspondence between $\mathbb{R}^{3n}$ and ordered unions of n squares is surjective and throughout this paper will serve as the domain for corresponding perimeter and area functions. As a general convention, when we refer to a figure $H \subset \mathbb{R}^{3n}$, we shall mean that H is the ordered union of n unit squares determined according to the correspondence described above. Define the function

$$\mathfrak{r} : \mathbb{R}^{3n} \to \mathbb{R}, \quad \mathfrak{r}(H) = \frac{p(H)}{\alpha(H)}.$$

That is, $\mathfrak{r}$ takes an ordered $3n$-tuple of identifiers as input and returns the ratio we've been examining for the figure identified by H. We'll refer to the vector $(\phi_1, \phi_2, \dots, \phi_n) \in \mathbb{R}^n$ as the *rotational displacement* of H. A figure $H \subset \mathbb{R}^2$ is said to have *distinct rotational displacement* if $\phi_i \neq \phi_j$ when $i \neq j$, is *vertex-free* if no vertex of H_i lies on the boundary of H_j whenever $i \neq j$, and is *triple-free* if no point lies on the boundaries of three distinct H_i. H is said to be in *standard position* provided:

(1) H has distinct rotational displacement,

(2) H is vertex-free, and

(3) H is triple-free.

The set of points in $\mathbb{R}^{3n}$ that do not have distinct rotational displacement lie on finitely many linear curves of the form $\phi_i = \phi_j + k\pi/2$, where $i \neq j$ and $k = 1, 2, 3$. Points which are not vertex-free lie on finitely many curves that are quadratic in the variables $\{s_i, t_i, \sin\phi_i, \cos\phi_i : i = 1, 2, \dots, n\}$, and points which are not triple-free lie on finitely many quartic curves in the same variables. It follows that the set of points which are in standard position is the complement of a *sparse set* in the sense that they are the complement of a countable union of monotonic curves and so are both residual and of full measure in $\mathbb{R}^{3n}$.

3. Continuity of perimeter and area

Here we give elementary geometric proofs that both the perimeter and area functions as we've defined them are continuous at configurations that are in standard position.

Lemma 1. *The perimeter function p is continuous at every point $H \in \mathbb{R}^{3n}$ which is in standard position.*

Proof. To show that perimeter is continuous, let $[a, b] \subset \operatorname{bd} H$ be a segment of maximal length on $\operatorname{bd} H$. As H has distinct rotational displacement there is a unique component square, say H_{i_o}, such that $[a, b] \subset \operatorname{bd} H_{i_o}$. To simplify notation, we assume $\phi_{i_o} = s_{i_o} = t_{i_o} = 0$, $a = (x_1, -1/2)$ and $b = (x_2, -1/2)$, where $-1/2 \leq x_1 < x_2 \leq 1/2$. We'll examine in some detail the case where neither a nor b are vertices of H_{i_o}; the other cases are similar.

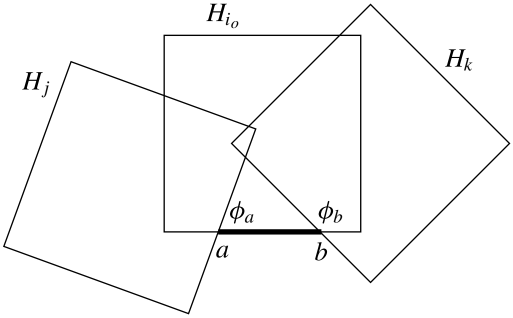


Figure 2. Since H is in standard position, a and b are uniquely determined by two additional component squares H_j and H_k.

Since H is in standard position, a and b are uniquely determined by two additional component squares H_j and H_k in the sense that $a = \operatorname{bd} H_{i_o} \cap \operatorname{bd} H_j$ and $b = \operatorname{bd} H_{i_o} \cap \operatorname{bd} H_k$. Let ϕ_a denote the angle determined by the intersection of the boundaries of H_{i_o} and H_j at a measured counterclockwise from the boundary of H_{i_o} to that of H_j; the angle ϕ_b is defined analogously. See Figure 2.

As $\phi_{i_o} = 0$, either $\phi_j = \phi_a$ or $\phi_j = \phi_a - \pi/2$. Also, $\phi_k = \phi_b$ or $\phi_k = \phi_b - \pi/2$. For definiteness we've supposed that $\phi_j = \phi_a$ and $\phi_k = \phi_b - \pi/2$ so that a is the intersection of the line ℓ given by $y = -1/2$ and ℓ_a given by $y = \tan\phi_a(x - x_1) + 1/2$. Similarly, b is the intersection of ℓ with the line ℓ_b given by $y = \tan\phi_b(x - x_2) + 1/2$. Using the fact that H is in standard position, there is an $\epsilon > 0$ such that $B_\epsilon([a, b])$ intersects no component square of H except H_{i_o}, H_j, and H_k.

Our immediate aim is to show that small perturbations of H result in small local perturbations of $\operatorname{bd} H$. To this end, suppose $\delta > 0$ and $d = (u_i, v_i, w_i)_{i=1}^n$ is a unit vector in $\mathbb{R}^{3n}$. Let $H^* = H + \delta \cdot d$ and denote its component squares by $H_i^* = H_i + \delta \cdot (u_i, v_i, w_i)$. Then, for δ sufficiently small, $B_\epsilon([a, b]) \cap H_i^* \neq \varnothing$ if and only if $i = i_o$, j or k. Let

(1) ℓ^* denote the line ℓ rotated by w_{i_o} about the center of H_{i_o}, $(0, 0)$, then translated by (u_{i_0}, v_{i_0}),

(2) ℓ_a^* denote the line ℓ_a rotated by w_j about the center of H_j, (s_j, t_j), then translated by (u_j, v_j),

(3) ℓ_b^* denote the line ℓ_b rotated by w_k about the center of H_k, (s_k, t_k), then translated by (u_k, v_k).

Finally, let $a^* = \ell^* \cap \ell_a^*$ and $b^* = \ell^* \cap \ell_b^*$. Then $[a^*, b^*]$ is a maximal segment on $\operatorname{bd} H^*$ and is the sole portion of $\operatorname{bd} H^*$ in $B_\epsilon([a, b])$. An elementary estimate shows that $\big| |[a^*, b^*]| - |[a, b]| \big| < 6\delta$.

As there are only finitely many such segments $[a, b] \subset \operatorname{bd} H_i$ comprising the boundary of H, and for δ sufficiently small, there is a one-to-one correspondence between these segments and those comprising the boundary of H^*, it follows that p is continuous at each point of standard position. □

The actual situation is that the perimeter function is continuous at a much larger set of points than those in standard position. The proof given above can be easily adapted to show that p is continuous at points having distinct rotational displacement; however, p is also continuous at most points that do not have distinct rotational displacement. Typical of points at which p is discontinuous is the point $H = (0, 0, 0, 1, .5, 0)$, where the perimeter is 7. If $H_n = (0, 0, 0, 1 + 1/n, .5, 1/n)$, then for every $n \in \mathbb{N}$, $p(H_n) = 8$ and yet $\{H_n\} \to H$.

A bit more can be said about the continuity of p at all points whether in standard position or not.

Proposition 2. *The function $p : \mathbb{R}^{3n} \to \mathbb{R}$ is lower semicontinuous.*

Proof. To see this, suppose $H = \bigcup_{i=1}^{n} H_i$ is an arbitrary configuration with component squares H_i. A segment $[a, b] \subset \operatorname{bd} H$ is called *proper* if $[a, b]$ is of maximal length under the restriction that no vertex of a component square of H lies on $[a, b]\backslash\{a, b\}$. The fact that H is the finite union of squares means that the boundary of H can be uniquely written as the nonoverlapping union of proper boundary segments. Suppose now that $.1 > \epsilon > 0$ is given and that $S = [a, b]$ is any proper segment on the boundary of H. Let $a^*, b^* \in S$ such that both $|a - a^*| = \epsilon|b - a|$ and $|b - b^*| = \epsilon|b - a|$ and set $S^* = [a^*, b^*]$. Let U be a ball about S^* such that $U \cap \operatorname{bd} H \subset S$ and the radius of the ball is less than $\epsilon/2$. For small ΔH, we wish to use $p(H)$ to estimate $p(H + \Delta H)$. First note that if S lies on a unique component square, then $S^* + \Delta H \subset (S + \Delta H) \cap U \subset \operatorname{bd}(H + \Delta H)$, and if all proper boundary segments have this property, we obtain an easy estimate of $p(H + \Delta H)$. However, should S be common to several component squares of H, then $S + \Delta H$ is the union of several segments and $\operatorname{bd}(H + \Delta H) \cap U$ is a piecewise linear selection from $S + \Delta H$. We handle this situation as follows. Let N_{a^*} denote the line segment in U that contains a^* and is normal to S; N_{b^*} is defined analogously for b^*. Let $a^{**} = (a + a^*)/2$, $b^{**} = (b + b^*)/2$ and $S^{**} = [a^{**}, b^{**}]$. We may take ΔH sufficiently small so that:

(1) $S^{**} + \Delta H \subset U$,

(2) $(S^{**} + \Delta H) \cap N_{a^*} \neq \varnothing \neq (S^{**} + \Delta H) \cap N_{b^*}$, and

(3) if T^{**} is analogous to S^{**}, but derived from another proper boundary segment, then $(T^{**} + \Delta H) \cap U = \varnothing$.

These conditions imply that a proper boundary segment for H yields a portion, but not necessarily a segment, of the boundary of $H + \Delta H$ that extends from N_{a^*} to N_{b^*}. As such, its length is at least $(1 - 2\epsilon)|b - a|$. Moreover, it follows from (3) above that distinct proper boundary segments of H yield disjoint boundary portions of $H+\Delta H$. Hence,

$$p(\text{bd}\, H+\Delta H)\geq p(\text{bd}\, H)(1-2\epsilon).$$

Since ϵ is arbitrary, it follows that

$$\liminf_{\Delta H\to 0} p(\text{bd}\, H+\Delta H)\geq p(\text{bd}\, H),$$

or that $p:\mathbb{R}^{3n}\to\mathbb{R}$ is lower semicontinuous. □

The minimizer for p is 4, occurring when all component squares of H coincide. The fact that p is lower semicontinuous, coupled with the continuity of the area function α, implies that the ratio p/α is lower semicontinuous and so has a minimizer. Establishing the minimizer for the ratio p/α and minimizers of similar configurations is interesting, but uses completely different methods from those of the current paper and is the topic of a separate study. It is not known if a maximizer of p/α exists.

We turn now to consider the area function.

Lemma 3. *The area function α is continuous at every point in $\mathbb{R}^{3n}$.*

Proof. As the area of each component square H_i is 1, it follows that the area function $\alpha:\mathbb{R}^3\to\mathbb{R}$ is Lipschitz in each coordinate with a Lipschitz constant of 1. Hence, α itself is Lipschitz with Lipschitz constant $\sqrt{3n}$. □

The following theorem now follows immediately from Lemmas 1 and 3.

Theorem 4. *The function $\mathfrak{r}$ is continuous at every point $H\in\mathbb{R}^{3n}$ which is in standard position.*

4. A derivative computation for perimeter

Next, we investigate the differentiability of the perimeter and area functions. Our goal is to prove the following theorem.

Theorem 5. *The perimeter function $p:\mathbb{R}^{3n}\to\mathbb{R}^+$ is differentiable at every point $H\in\mathbb{R}^{3n}$ in standard position.*

Proof. We show the partial derivatives of p exist and are continuous at each point of standard position. Fix $0\leq i_o\leq n$ and consider the three partial derivatives $\partial p/\partial s_{i_o}$, $\partial p/\partial t_{i_o}$ and, initially, $\partial p/\partial \phi_{i_o}$.

Part 1: $\partial p/\partial\phi_{i_o}$. As in Lemma 1, we take a particular component segment $[a,b]$ on the boundary of H and again note that $[a,b]$ must lie on the boundary of a single H_j. There are two cases depending on whether $H_j=H_{i_o}$.

Case 1a: $[a,b]$ lies on the boundary of H_{i_o}. We adopt the notation of Lemma 1 in its entirety so that $\phi_{i_o}=s_{i_o}=t_{i_o}=0$, $a=(x_1,-1/2)$ and $b=(x_2,-1/2)$, where $-1/2\leq x_1<x_2\leq 1/2$. The lines ℓ, ℓ_a and ℓ_b are also as before. However, the unit

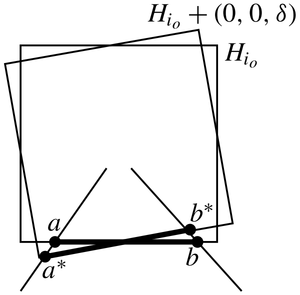


Figure 3. $[a, b]$ and $[a^*, b^*]$ in the case that $[a, b]$ lies on the boundary of H_{i_o}.

vector we consider is more specific in this case; $d \in \mathbb{R}^{3n}$ is the vector with $w_{i_o} = 1$ and all remaining components 0. The set $H^* = H + \delta \cdot d$ is comprised of precisely the same component squares as H with the exception of H_{i_o}, which is replaced by $H_{i_o} + (0, 0, \delta)$, a rotation of H_{i_o} about its center by an angle of δ. The segment on $\operatorname{bd} H^*$ corresponding to $[a, b]$ is $[a^*, b^*]$, where $a* = \ell^* \cap \ell_a$ and $b^* = \ell^* \cap \ell_b*$ since $\ell_a = \ell_a^*$ and $\ell_b = \ell_b^*$. See Figure 3.

A computation yields

$$\begin{aligned} a^* &= \left(\frac{x_1 \tan\phi_a}{\tan\phi_a - \tan\delta}, \frac{x_1 \tan\phi_a \tan\delta}{\tan\phi_a - \tan\delta} - \frac{1}{2}\right), \\ b^* &= \left(\frac{x_2 \tan\phi_b}{\tan\phi_b + \tan\delta}, \frac{x_2 \tan\phi_b \tan\delta}{\tan\phi_b + \tan\delta} - \frac{1}{2}\right). \end{aligned} \tag{1}$$

Consequently, the x-coordinate of $b^* - a^*$ is

$$x(b^* - a^*) = \frac{(x_2 - x_1)\tan\phi_a \tan\phi_b - \tan\delta(x_2 \tan\phi_b + x_1 \tan\phi_a)}{(\tan\phi_a - \tan\delta)(\tan\phi_b + \tan\delta)}.$$

However, $x(b^* - a^*)/|b^* - a^*| = \cos\delta$ and so

$$|b^* - a^*| = \frac{(x_2 - x_1)\tan\phi_a \tan\phi_b - \tan\delta(x_2 \tan\phi_b + x_1 \tan\phi_a)}{(\tan\phi_a - \tan\delta)(\tan\phi_b + \tan\delta)\cos\delta}. \tag{2}$$

We're now in a position to complete the computation of the contribution of $[a, b]$ to $\partial p/\partial\phi_{i_o}$ at H_{i_o}:

$$\lim_{\delta\to 0} \frac{|b^* - a^*| - |b - a|}{\delta} = (x_2 - x_1)(\cot\phi_b - \cot\phi_a). \tag{3}$$

Case 1b: $[a, b]$ lies on the boundary of H_j with $i_o \neq j$. If $[a, b] \cap H_{i_o} = \varnothing$, then the partial derivative of that portion of p with respect to ϕ_{i_o} is 0. Hence, we may assume that $[a, b] \cap H_{i_o} \neq \varnothing$. As $[a, b] \subset \operatorname{bd} H$ is maximal, it follows that either $[a, b] \cap H_{i_o} = \{a\}$ or $[a, b] \cap H_{i_o} = \{b\}$. We suppose the former and for purposes

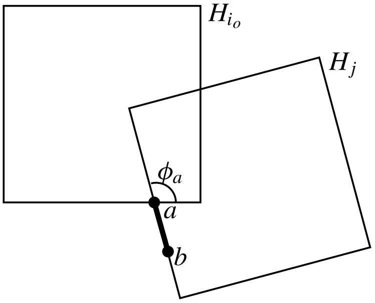


Figure 4. $[a, b] \not\subset \operatorname{bd} H_{i_o}$.

of computation, we again adopt some of the notation of Lemma 1. Specifically we suppose that $H_{i_o} = (0, 0, 0)$, $a = (x_1, -1/2)$, ϕ_a, ℓ_a, ℓ and ℓ^* are as before. Then the segment $[a, b]$ lies on the line ℓ_a. See Figure 4 where a lies to the right of the center of H_{i_o} (or $0 \le x_1 \le 1/2$).

In this case, the change in perimeter due to $[a, b]$ is $\big||[a^*, b^*]| - |[a, b]|\big| = |a^* - a|$. Then according to the law of sines,

$$|a^* - a| = \frac{|c - a| \sin\delta}{\sin(\phi_\delta)} = \frac{|c - a| \sin\delta}{\sin\phi_a \cos\delta - \cos\phi_a \sin\delta},$$

where $c = \ell \cap \ell^*$. See Figure 5.

Hence, in this case, the contribution of $[a, b]$ to $\partial p/\partial\phi_{i_o}$ at H_{i_o} is

$$\lim_{\delta\to 0} \frac{|a^* - a|}{\delta} = \lim_{\delta\to 0} \frac{|c - a|(\sin\delta)/\delta}{\sin\phi_a \cos\delta - \cos\phi_a \sin\delta} = \frac{x_1}{\sin\phi_a}.$$

Since H is in standard position, this is well-defined and continuous. The case in which a lies to the left of the center of H_{i_o} is the same, but is negative since $-1/2 \le x_1 < 0$ in that case.

To summarize, we have shown that the rate of change of each component segment of $\operatorname{bd} H$ with respect to ϕ_{i_o} is continuous. As $\operatorname{bd} H$ is comprised of finitely many

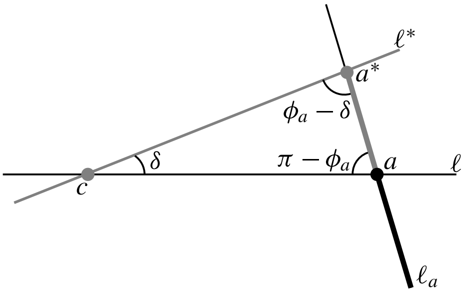


Figure 5. $\big||[a^*, b^*]| - |[a, b]|\big| = |a^* - a|$ in the case that $[a, b] \not\subset \operatorname{bd} H_{i_o}$.

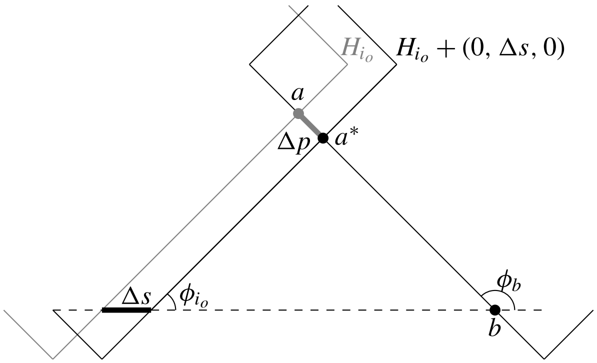


Figure 6. Translating H_{i_o} by $(0, \Delta s, 0)$.

such segments, it follows that $\partial p/\partial\phi_{i_o}$ exists and is continuous at each point in standard position.

Part 2: $\partial p/\partial s_i$. For notational convenience, we denote Δs_{i_o} simply by Δs. As in the previous case, we take a particular segment $[a, b]$ on the boundary of H and consider three cases:

(a) $[a, b]$ does not lie on bd H_{i_o},

(b) $[a, b]$ lies on bd H_{i_o} and contains a vertex of H_{i_o}, and

(c) $[a, b]$ lies on bd H_{i_o} and neither a nor b are vertices of H_{i_o}.

Case 2a: Suppose that $[a, b] \not\subset$ bd H_{i_o}. If $[a, b] \cap H_{i_o} = \varnothing$, a sufficiently small translation of H_{i_o} will leave $[a, b]$ unchanged. Suppose then that $[a, b] \cap H_{i_o} \neq \varnothing$. As H is in standard position, it follows that either $[a, b] \cap H_{i_o} = \{a\}$ or $[a, b] \cap H_{i_o} = \{b\}$. Suppose bd $H_i \cap [a, b] = \{a\}$. If Δs is sufficiently small, the point b remains on bd H when translating H_{i_o} by Δs. Because a lies on the intersection of two component squares and H is in standard position, a is not the vertex of any square. Hence, $\mathrm{bd}(H_{i_o} + (0, \Delta s, 0)) \cap [a, b] \neq \varnothing$. Let $\mathrm{bd}(H_{i_o} + (0, \Delta s, 0)) \cap [a, b] = a^*$. The segments $[a^*, b]$ and $[a, b]$ differ by a length of Δp and it is this distance we wish to compute. See Figure 6.

We consider two cases depending on if $\phi_{i_o} < \phi_b$ or $\phi_{i_o} > \phi_b$. Supposing that $\phi_{i_o} < \phi_b$, the relevant triangle is illustrated in Figure 7.

Using the law of sines, we compute $\Delta p/\Delta s = \sin\phi_{i_o}/\sin(\phi_b - \phi_{i_o})$. If $\phi_{i_o} > \phi_b$, a similar computation yields $\Delta p/\Delta s = -\sin\phi_{i_o}/\sin(\phi_b - \phi_{i_o})$. As H is in standard position, $\phi_{i_o} \neq \phi_b$, so these are the only cases.

Case 2b: Suppose that $[a, b] \subset$ bd H_{i_o} and that a is a vertex of H_{i_o}. If b is also a vertex of H_{i_o}, then as H is in standard position, neither a nor b lie on the boundary of another component square. Consequently, $[a+\Delta s, b+\Delta s] \subset \mathrm{bd}(H+(0, \Delta s, 0))$, given a sufficiently small translation Δs.

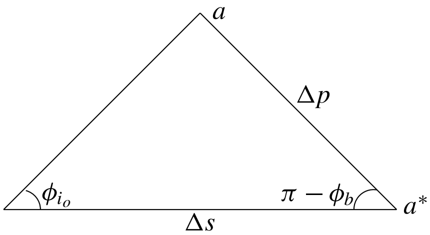


Figure 7. The relevant triangle in the case that $\phi_{i_o} < \phi_b$.

Suppose then that b is not a vertex of H_{i_o}. Then, given a sufficiently small translation Δs, the segment $[a + \Delta s, b + \Delta s]$ will intersect bd H at point b^* and $[b, b^*] \subset \text{bd}\, H$. See Figure 8.

At this point the geometry and subsequent computation of $\Delta p/\Delta s$ are analogous to that found in Figure 7. In the case illustrated in Figure 8, $\Delta p/\Delta s = \cos\phi_{i_o} - \sin\phi_{i_o}\tan\phi_b$.

Case 2c: Finally, suppose that $[a, b]$ lies on bd H_{i_o} and neither a nor b are vertices of H_{i_o}. Then the endpoints a and b of this boundary segment lie on uniquely determined segments on bd H that lie on the boundaries of component squares other than H_{i_o}. This is similar to the analysis done in Case 1a above. As H is in standard position, if Δs is sufficiently small, then as H_{i_o} is translated to $H_{i_o} + (0, \Delta s, 0)$, the boundary segment $[a, b]$ is translated to a new boundary segment $[a^*, b^*]$. Moreover, a^* lies on the same boundary segment of H as does a, and b^* lies on the same boundary segment of H as does b. See Figure 9.

In order to facilitate the required computation, we establish the notation found in Figure 10.

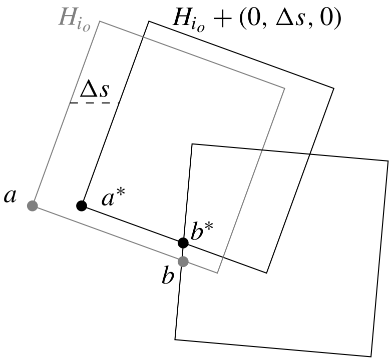


Figure 8. The case in which b is not a vertex of H_{i_o}.

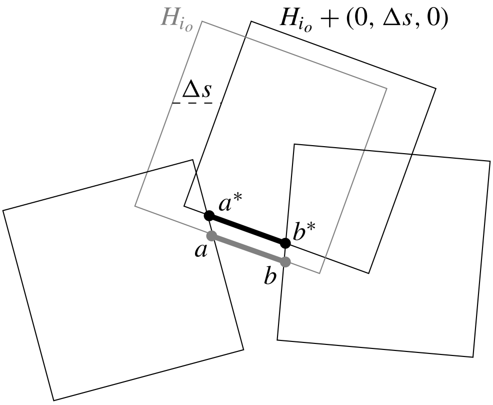


Figure 9. The case in which $[a, b]$ lies on bd H_{i_o} and neither a nor b are vertices of H_{i_o}.

Applying the law of sines to triangles $[a, a^*, c]$ and $[a, a^*, d]$, we find

$$\frac{|a-a^*|}{\sin\phi} = \frac{\Delta s}{\sin(\pi-\phi-\phi_{i_o})},$$
$$\frac{|a-a^*|}{\sin(\phi+\phi_b)} = \frac{\Delta p}{\sin(\phi_{i_o}-\phi_b)}.$$

Hence,

$$\frac{\Delta p}{\Delta s} = \frac{\sin\phi\,\sin(\phi_{i_o}-\phi_b)}{\sin(\phi+\phi_{i_o})\,\sin(\phi+\phi_b)}.$$

In the case illustrated in Figure 10, $\phi = \pi/2 - \phi_{i_o}$. As H is in standard position, this quantity is well-defined and indeed constant.

To obtain $\partial p/\partial s_i$, sum $\Delta p/\Delta s$ for every component segment $[a, b] \subset$ bd H. Since there are finitely many partitioning segments and $\Delta p/\Delta s$ is continuous for each of them, $\partial p/\partial s_i$ exists and is continuous at every point in standard position.

Part 3: $\partial p/\partial y_i$. Rotating the whole figure by 90°, this case becomes exactly the same as the previous one.

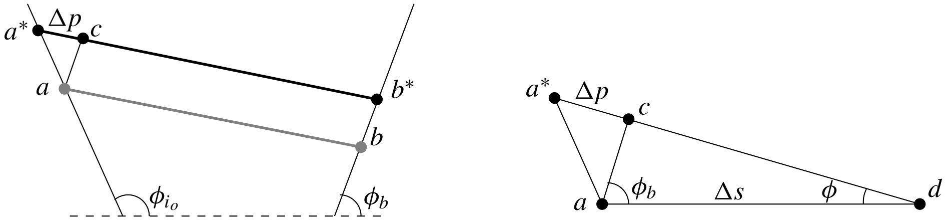


Figure 10. The same case with added detail and notation.

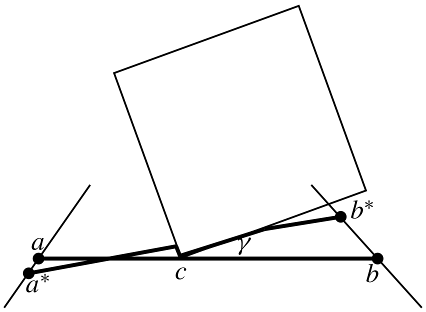


Figure 11. When the segment $[a, b]$ contains the vertex of a new square, the derivative of perimeter changes discontinuously.

Finally, because all of the partial derivatives exist and are continuous at each point in standard position, the perimeter function is differentiable at every point $H \in \mathbb{R}^{3n}$ of standard position, as desired. □

The situation for points that are not in standard position is mixed. For example, at some of these points the perimeter is differentiable; if the configuration H is not vertex-free, but the vertices that lie on edges of remote squares are all interior to H, then the fact that H is not vertex-free has no bearing on the differentiability of perimeter at H. However, if a segment on the boundary of H contains a vertex, then p is not differentiable at H. Figure 11 is a portion of Figure 3 with an additional square added in such a way that a new vertex, c resides on the segment $[a, b]$. The angle this new square makes with the segment $[a, b]$ is important and labeled γ. Also important is the distance $|c-a|$ this vertex is from the endpoint a. In the next paragraph, we adopt the notation established earlier in Lemma 1 and Part 1 in the proof of Theorem 5.

In such a case, several of the partial derivatives do not exist. In particular, both one-sided partial derivatives, $\partial p/\partial\phi_{i_o}^+$ and $\partial p/\partial\phi_{i_o}^-$ exist, but differ. The latter, $\partial p/\partial\phi_{i_o}^-$ is as computed in Part 1 of the proof of Theorem 5. However, a computation similar to this shows that $\partial p/\partial\phi_{i_o}^+$ differs from $\partial p/\partial\phi_{i_o}^-$ by $|c-a|\tan(\gamma)$. See Figure 11. To summarize,

$$\frac{\partial p}{\partial s_{i_o}^-} = (x_2 - x_1)(\cot\phi_b - \cot\phi_a),$$

$$\frac{\partial p}{\partial s_{i_o}^+} = (x_2 - x_1)(\cot\phi_b - \cot\phi_a) + |c-a|\tan\alpha.$$

5. A derivative computation for area

To show $\mathfrak{r}$ is differentiable at every point in standard position, it remains to show that the area function is differentiable at every point in standard position and to show how that derivative can be computed.

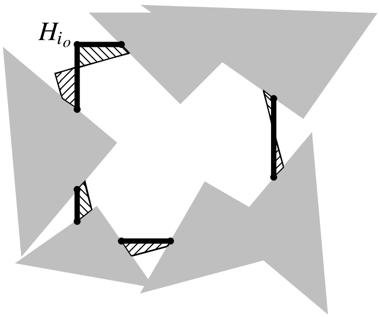


Figure 12. The change in area after rotating H_{i_o}.

Theorem 6. *The area function $\alpha : \mathbb{R}^{3n} \to \mathbb{R}^+$ is differentiable at every point $H \in \mathbb{R}^{3n}$ in standard position.*

Proof. Again we show the partial derivatives of α exist and are continuous at each point of standard position. Fix $0 \le i_o \le n$. If $H_{i_o} \subset \operatorname{int}(H)$, then $\partial a/\partial s_{i_o}(H) = \partial a/\partial t_{i_o}(H) = \partial a/\partial \phi_{i_o}(H) = 0$, so we may assume that a portion of $\operatorname{bd} H_{i_o}$ is contained in $\operatorname{bd} H$; since H is in standard position, $\operatorname{bd} H_{i_o} \cap \operatorname{bd} H$ is the union of closed nonoverlapping intervals.

Part 1: $\partial\alpha/\partial\phi_{i_o}$. There may be several segments common to $\operatorname{bd} H_{i_o}$ and $\operatorname{bd} H$, and for a fixed $\delta = \Delta\phi_{i_o}$, each such segment contributes to a corresponding $\Delta\alpha$. See Figure 12 where those segments common to both $\operatorname{bd} H_{i_o}$ and $\operatorname{bd} H$ are darkened and the components of $\Delta\alpha$ are hatched, northeast for gain and northwest for loss. The gray regions are portions of the other component squares of H that intersect H_{i_o}.

The area change, $\Delta\alpha$ is simply the sum total of the signed area changes at each of the line segment components of $\operatorname{bd} H_{i_o} \cap \operatorname{bd} H$. There are several cases to consider depending on the relative location of a boundary segment of $\operatorname{bd} H_{i_o} \cap \operatorname{bd} H$, but in any case, for purposes of this computation, we may assume $H_{i_o} = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ and that the boundary segment $[a, b]$ lies on the line $y = -\frac{1}{2}$.

Case 1a: $a = \left(x_1, -\frac{1}{2}\right)$ and $b = \left(x_2, -\frac{1}{2}\right)$ with $-\frac{1}{2} < x_1 < x_2 \le 0$. Then the additional area determined by $[a, b]$ is a net gain (or +) and is the area of the quadrilateral $[a, p, q, b]$. See Figure 13 where the notation is the same as in the Lemma 1 except for a new value of $d = (0, \dots, 0, \Delta\phi_{i_o}, 0, \dots, 0)$. Using the coordinates of a^* and b^* computed in (1), we find the area of the quadrilateral $[a, a^*, b^*, b]$ to be

$$\Delta\alpha([a,b]) = \frac{(x_1^2 - x_2^2)\tan\phi_a \tan\phi_b \tan\delta + \tan^2\delta\,(x_2^2 \tan\phi_b + x_1^2 \tan\phi_a)}{2(\tan\phi_a - \tan\delta)(\tan\phi_b + \tan\delta)}. \tag{4}$$

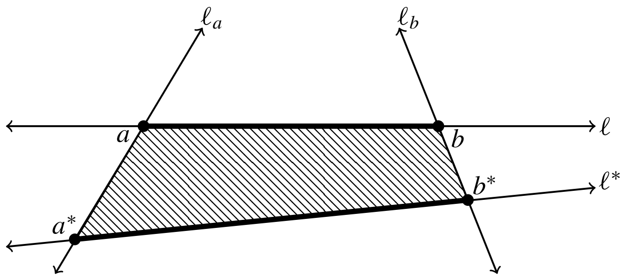


Figure 13. $\Delta\alpha$ at $[a, b]$ in Case 1a.

To find the contribution to $\partial\alpha/\phi_{i_o}$ at $[a, b]$, we divide the quantity found in (4) by δ and take the limit as $\delta \to 0$ to obtain

$$\left.\frac{\partial\alpha}{\phi_{i_o}}\right|_{[a,b]} = \lim_{\delta\to 0} \frac{\Delta\alpha([a, b])}{\delta} = \frac{x_1^2 - x_2^2}{2}.$$

Case 1b: $a = \left(x_1, -\frac{1}{2}\right)$ and $b = \left(x_2, -\frac{1}{2}\right)$ with $0 \le x_1 < x_2 \le \frac{1}{2}$. This case is symmetric to Case 1a, but since $0 \le x_1 < x_2$, the contribution to $\partial\alpha/\phi_{i_o}$ is negative:

$$\left.\frac{\partial\alpha}{\phi_{i_o}}\right|_{[a,b]} = \frac{x_1^2 - x_2^2}{2}.$$

Case 1c: $a = \left(x_1, -\frac{1}{2}\right)$ and $b = \left(x_2, -\frac{1}{2}\right)$ with $-\frac{1}{2} \le x_1 \le 0 \le x_2 \le \frac{1}{2}$. Once again

$$\left.\frac{\partial\alpha}{\phi_{i_o}}\right|_{[a,b]} = \frac{x_1^2 - x_2^2}{2}.$$

To see this, introduce $x_3 = 0$ and add the corresponding amounts computed using the formula from Cases 1a and 1b.

Part 2: $\partial\alpha/\partial s_{i_o}$. The analysis of this case is much the same as Part 1. Again there may be several segments common to bd H_{i_o} and bd H, and for a fixed $\delta = \Delta s_{i_o}$, each such segment contributes to a corresponding $\Delta\alpha$. See Figure 14 where those segments common to both bd H_{i_o} and bd H are darkened and the components of $\Delta\alpha$ are hatched, northeast for gain and northwest for loss. The gray regions are portions of the other component squares of H that intersect H_{i_o}.

There are two basic cases to consider here. The first concerns a component segment $[a, b] \subset \text{bd}\, H_{i_o} \cap \text{bd}\, H$ that lies on either the bottom or top of H_{i_o} and the second is when that segment lies on one of the other two sides. In both cases we let $\Delta s_{i_o} = \Delta s$ be sufficiently small.

Case 2a: $[a, b]$ lies on the bottom of H_{i_o}. For definiteness, we again suppose that neither a nor b is a vertex of H_{i_o}, as the cases when they are vertices can be handled

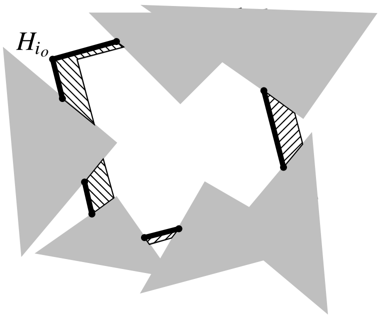


Figure 14. The change in area after translating H_{i_o}.

in much the same manner. As before, there are uniquely defined H_j and H_k such that $a \in \text{bd}\, H_j \cap \text{bd}\, H_{i_o}$ and $b \in \text{bd}\, H_k \cap \text{bd}\, H_{i_o}$. Also let ℓ denote the line containing $[a, b]$, so that the segment $[a, b]$ is that portion of ℓ between H_j and H_k. Similarly, let ℓ^* be the line $\ell + (0, \Delta s, 0)$, and let $[a^*, b^*]$ denote that segment on ℓ^* extending between H_j and H_k. See Figure 15 where the trapezoidal region whose area is $\Delta\alpha$ at $[a, b]$ is shaded and the rotation displacement ϕ_{i_o} as well as Δs are labeled.

Then

$$\Delta\alpha|_{[a,b]} = \frac{|b-a| + |b^* - a^*|}{2} \sin\phi_{i_o} \cdot \Delta s.$$

From this it follows that

$$\left.\frac{\partial\alpha}{\partial s_{i_o}}\right|_{[a,b]} = \lim_{\Delta s \to 0} \frac{|b-a| + |b^* - a^*|}{2} \sin\phi_{i_o} = |b-a| \sin\phi_{i_o}.$$

The case in which $[a, b]$ lies on the top of H_{i_o} is identical with the exception that the sign is negative.

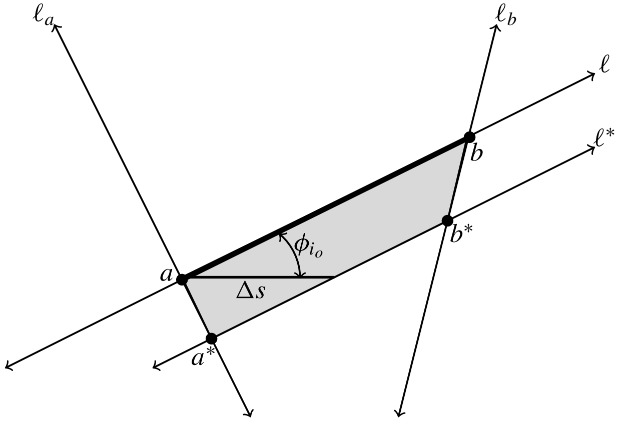


Figure 15. $\Delta\alpha$ at $[a, b]$ in Case 2a.

Case 2b: $[a, b]$ lies on the right (or left) side of H_{i_o}. The right side case is much the same as described in Case 2a above, but with the angle ϕ_{i_o} replaced by $\phi_{i_o} + \pi/2$. The resulting derivative formula becomes

$$\frac{\partial\alpha}{\partial s_{i_o}}\bigg|_{[a,b]} = \lim_{\Delta s\to 0} \frac{|b-a| + |b^* - a^*|}{2} \cos\phi_{i_o} = |b-a|\cos\phi_{i_o}.$$

As above, the left side case is identical with the exception that the sign is negative.

Part 3: $\partial\alpha/\partial t_{i_o}$. As with the analysis of perimeter, the translation cases are completely analogous.

As each of the partial derivatives is defined and continuous at each point H in standard position, the proof of Theorem 6 is complete. □

Because the perimeter and area functions are differentiable at every point in standard position (and $a \neq 0$), their ratio $\mathfrak{r} : \mathbb{R}^{3n} \to \mathbb{R}^+$ is differentiable and our main result now follows immediately.

Theorem 7. *The function $\mathfrak{r} : \mathbb{R}^{3n} \to \mathbb{R}^+$ is differentiable at every point $H \in \mathbb{R}^{3n}$ in standard position.*

The idea of studying the variation of $p(H)/a(H)$ allows us to consider a vast body of discrete geometric literature to help study the problem. However, nearly all of this literature concerns itself with studying disks in the plane, rather than squares or arbitrary shapes. An example is the famous Kneser–Poulsen theorem concerning disks in the plane. See [Bezdek and Connelly 2002] and [Bollobás 1968] for details.

Kneser–Poulsen theorem. *If a set of disks in the plane are rearranged so that the distance between the centers of any pair of discs decreases, then the area and the perimeter of the union of the discs also decreases.*

References

[Bezdek and Connelly 2002] K. Bezdek and R. Connelly, "Pushing disks apart: the Kneser–Poulsen conjecture in the plane", *J. Reine Ang. Math.* **553** (2002), 221–236. MR 2003m:52001 Zbl 1021.52012

[Bollobás 1968] B. Bollobás, "Area of the union of disks", *Elem. Math.* **23** (1968), 60–61. MR 38 #3772 Zbl 0153.51903

[Cheng and Edelsbrunner 2003] H.-L. Cheng and H. Edelsbrunner, "Area and perimeter derivatives of a union of disks", pp. 88–97 in *Computer science in perspective: essays dedicated to Thomas Ottmann*, edited by R. Klein et al., Lecture Notes in Computer Science **2598**, Springer, New York, 2003. Zbl 1023.68108

[Competition 1998] Anonymous, "Schweitzer Miklós Matematikai Emlékverseny", 1998, available at www.math.u-szeged.hu/~mmaroti/schweitzer/schweitzer-1998.pdf. Problems from the Miklós Schweitzer Memorial Mathematical Competition.

[Gyenes 2005] Z. Gyenes, *The ratio of the surface-area and volume of finite union of copies of a fixed set in $\mathbb{R}^n$*, master's thesis, Eötvös Loránd University, Budapest, 2005, available at http://www.cs.elte.hu/~dom/z.pdf.

[Humke et al. 2015] P. D. Humke, C. Marcott, B. Mellem, and C. Stiegler, "Bounded – yes, but 4?", preprint, 2015. arXiv 1507.08536

[Keleti 1998] T. Keleti, "A covering property of some classes of sets in $\mathbb{R}^2$", *Acta Univ. Carolin. Math. Phys.* **39**:1-2 (1998), 111–118. MR 2000g:28003 Zbl 1016.28003

Received: 2014-10-02 Revised: 2014-11-19 Accepted: 2014-11-20

humkep@gmail.com
Department of Mathematics, St. Olaf College,
1520 St. Olaf Avenue, Northfield, MN 55057, United States

cam.marcott@gmail.com
Department of Mathematics, St. Olaf College,
1520 St. Olaf Avenue, Northfield, MN 55057, United States

contactatrius@hotmail.com
Department of Mathematics, St. Olaf College,
1520 St. Olaf Avenue, Northfield, MN 55057, United States

cole.stiegler@gmail.com
Department of Mathematics, St. Olaf College,
1520 St. Olaf Avenue, Northfield, MN 55057, United States

INVOLVE 8:5 (2015)
dx.doi.org/10.2140/involve.2015.8.893

On the Levi graph of point-line configurations

Jessica Hauschild, Jazmin Ortiz and Oscar Vega

(Communicated by Joseph A. Gallian)

We prove that the well-covered dimension of the Levi graph of a point-line configuration with v points, b lines, r lines incident with each point, and every line containing k points is equal to 0, whenever $r > 2$.

1. Introduction

The concept of the well-covered space of a graph was first introduced by Caro, Ellingham, Ramey, and Yuster [Caro et al. 1998; Caro and Yuster 1999] as an effort to generalize the study of well-covered graphs. Brown and Nowakowski [2005] continued the study of this object and, among other things, provided several examples of graphs featuring odd behaviors regarding their well-covered spaces. One of these special situations occurs when the well-covered space of the graph is trivial, i.e., when the graph is *anti-well-covered*. In this work, we prove that almost all Levi graphs of configurations in the family of the so-called (v_r, b_k)-configurations (see Definition 3) are anti-well-covered.

We start our exposition by providing the following definitions and previously known results. Any introductory concepts we do not present here may be found in the books by Bondy and Murty [1976] and Grünbaum [2009].

We consider only simple and undirected graphs. A graph will be denoted by $G = (V(G), E(G))$, as is customary, where $V(G)$ is the set of vertices of the graph and $E(G)$ is the set of edges of the graph. We think of $E(G)$ as an irreflexive symmetric relation on $V(G)$. Two vertices of a graph are said to be *adjacent* if they are connected by an edge. An *independent* set of vertices is one in which no two vertices in the set are adjacent. If an independent set, M, of a graph G is not a proper subset of any other independent set of G, then M is a *maximal independent set* of G.

Definition 1. Let G be a graph and F a field.

(1) A function $f : V(G) \to F$ is said to be a *weighting* of G. If the sum of all weights is constant for all maximal independent sets of G, then the weighting is a *well-covered weighting* of G.

MSC2010: primary 05B30; secondary 51E05, 51E30.
Keywords: Levi graph, maximal independent sets, configurations.

(2) The F-vector space consisting of all well-covered weightings of G is called the well-covered space of G (relative to F).

(3) The dimension of this vector space is called the *well-covered dimension* of G, denoted $\operatorname{wcdim}(G, F)$.

Remark 1. For some graphs, the characteristic of the field F makes a difference when calculating the well-covered dimension (see [Birnbaum et al. 2014] and [Brown and Nowakowski 2005]). If $\operatorname{char}(F)$ does not cause a change in the well-covered dimension, then the well-covered dimension is denoted as $\operatorname{wcdim}(G)$.

In order to calculate the well-covered dimension of a graph, G, one would generally need to find all possible maximal independent sets of G. However, finding all maximal independent sets is not always an easy task, as this is a known NP-complete problem.

Despite the NP-complete nature of this problem, let us assume that we have found all possible maximal independent sets of G. We will denote these maximal independent sets as M_i for $i = 0, 1, \dots, k-1$. The well-covered weightings of G are determined by solving a system of linear equations that arise from considering all equations of the form $M_0 = M_i$ for $i = 1, \dots, k-1$. We replace this system with the equivalent homogeneous one via standard operations and create an associated matrix A_G. Observe that the dimension of the nullspace of A_G is equal to the dimension of the well-covered space of G. Thus,

$$\operatorname{wcdim}(G, F) = |V(G)| - \operatorname{rank}(A_G).$$

We now move onto another component of our work: configurations.

Definition 2. A (point-line) configuration is a triple $(\mathcal{P}, \mathcal{L}, \mathcal{I})$, where $\mathcal{P}$ is set of points, $\mathcal{L}$ is a set of lines, and $\mathcal{I}$ is an incidence relation between $\mathcal{P}$ and $\mathcal{L}$, that has the following properties:

(1) Any two points are incident with at most one line.

(2) Any two lines are incident with at most one point.

Next, there is some notation for configurations that needs to be set, as well as specific parameters that need to be established for the main result of this work.

Definition 3. We define a (v_r, b_k)-configuration as a configuration such that

(1) $|\mathcal{P}| = v$, and $v \geq 4$.

(2) $|\mathcal{L}| = b$, and $b \geq 4$.

(3) There are exactly k points incident with each line, and $k \geq 2$.

(4) There are exactly r lines incident with each point, and $r \geq 2$.

When $v = b$ and $r = k$, the configuration will be denoted by (v_r).

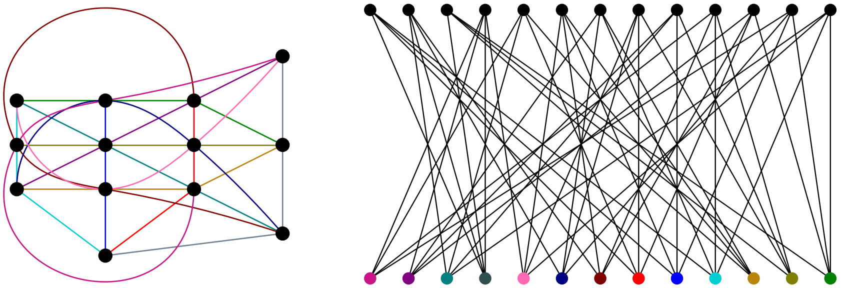


Figure 1. $(13_4) = PG(2,3)$ and $\text{Levi}_{(13_4)}$.

Example 1. Several well-known geometric structures fall into the category of (v_r, b_k)-configurations. For instance:

(1) A projective plane of order q is a $(q^2+q+1_{(q+1)})$-configuration, where q is the power of a prime. See Figure 1 for a representation of $PG(2,3) = (13_4)$.

(2) The Pappus configuration is a (9_3)-configuration, and the Desargues configuration is a (10_3)-configuration.

(3) $PG(n,q)$ is a
$$\left(\frac{q^{n+1}-1}{q-1}_{(q+1)}, \frac{(q^{n+1}-1)(q^n-1)}{(q^2-1)(q-1)}_{(q^2+q+1)}\right)\text{-configuration,}$$
where q is the power of a prime.

(4) A generalized quadrangle $G(s,t)$ is a $((1+s)(st+1)_{(1+s)}, (1+t)(st+1)_{(1+t)})$-configuration.

The reader is referred to the book by Batten [1997] for more information about these important geometric objects.

Finally, we define Levi graphs, which will connect configurations and graphs.

Definition 4. The Levi graph of a (v_r, b_k)-configuration $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ is the bipartite graph G with $V(G) = \mathcal{P} \cup \mathcal{L}$ and $E(G) = \mathcal{I}$. That is, $p \in \mathcal{P}$ is adjacent to $\ell \in \mathcal{L}$ if and only if $p\mathcal{I}\ell$. We will denote this graph $\text{Levi}_{(v_r,b_k)}$.
Note that $\mathcal{P}$ and $\mathcal{L}$ are independent sets — the partite sets — in G.

Our main result, which will be proven in the following section, combines all of these objects as follows:

Theorem 1. *If r is a positive integer greater than* 2, *then* $\text{wcdim}(\text{Levi}_{(v_r,b_k)}) = 0$.

We would like to remark that Theorem 1 says is that almost all Levi graphs of (v_r, b_k)-configurations are anti-well-covered.

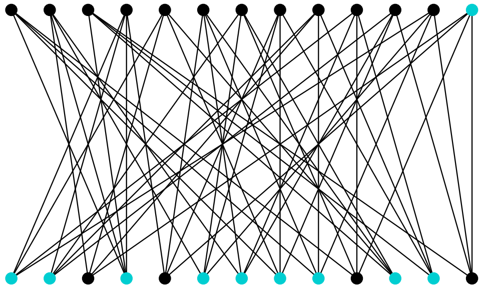


Figure 2. A maximal independent set M_P in $\text{Levi}_{(13_4)}$.

2. The well-covered dimension of $\text{Levi}_{(v_r,b_k)}$

We will prove Theorem 1 by first proving a technical lemma that introduces a family of maximal independent sets that will prove to be useful later on.

Lemma 1. *A Levi graph of a configuration* (v_r, b_k), *where* $r > 2$, *has at least* $v + b + 2$ *maximal independent sets.*

Proof. Let P be a fixed point in (v_r, b_k). We consider the set, M_P, of vertices of $\text{Levi}_{(v_r,b_k)}$ given by P and all the lines not incident to P. This is an independent set of $\text{Levi}_{(v_r,b_k)}$ because there is no incidence between vertices in the set. Moreover, note that if we included another point-vertex to M_P, then that vertex would be adjacent to one of the line-vertices in M_P (because of condition (2) in Definition 2, and the fact that $r > 2$). Also, if another line-vertex were to be added to M_P, then this line would have to be incident with P. It follows that M_P is a maximal independent set of $\text{Levi}_{(v_r,b_k)}$. See Figure 2 for an example.

Repeating this construction for all v points in (v_r, b_k), we get v distinct maximal independent sets of $\text{Levi}_{(v_r,b_k)}$.

We will now construct another b distinct maximal independent sets of $\text{Levi}_{(v_r,b_k)}$. We start by fixing a line ℓ in (v_r, b_k) and then any two distinct points $P_1, P_2 \in \ell$ (recall that $k \geq 2$). We consider the set, M_{P_1,P_2} of vertices of $\text{Levi}_{(v_r,b_k)}$ given by P_1, P_2 and all the lines not incident to either of these points. Note that this forms an independent set since adjacency in $\text{Levi}_{(v_r,b_k)}$ only occurs if incidence occurs in (v_r, b_k). If we try to add in another vertex-point to M_{P_1,P_2}, since $r > 2$, this point will be incident to one of the lines not through P_1 or P_2 and will therefore be adjacent to the vertex-lines in M_{P_1,P_2}. If we try to add another vertex-line to M_{P_1,P_2}, then this line will be incident to one or both of P_1 and P_2. Therefore, M_{P_1,P_2} is a maximal independent set of $\text{Levi}_{(v_r,b_k)}$. See Figure 3 for an example.

Repeating this construction for all b lines in (v_r, b_k) (it does not matter what pair of points one picks on any given line), we get b distinct maximal independent sets of $\text{Levi}_{(v_r,b_k)}$.

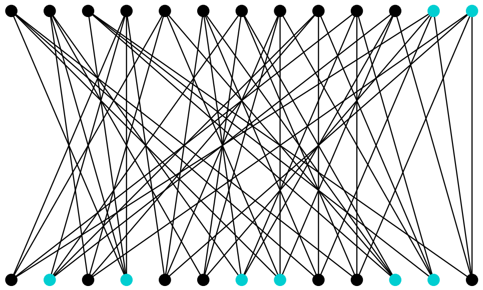


Figure 3. A maximal independent set M_{P_1,P_2} in $\mathrm{Levi}_{(13_4)}$.

Finally, note that the set of all point-vertices in $\mathrm{Levi}_{(v_r,b_k)}$ is a maximal independent set and the set of all line-vertices in $\mathrm{Levi}_{(v_r,b_k)}$ is as well. Hence, we have constructed $v+b+2$ distinct maximal independent sets in $\mathrm{Levi}_{(v_r,b_k)}$. □

Next, we proceed to prove our main result.

Proof of Theorem 1. We denote by F the field of scalars of the well-covered space of $G = \mathrm{Levi}_{(v_r,b_k)}$, where $r > 2$. Let A_G be the associated matrix of G, and note that A_G has $v+b$ columns. In order to prove that A_G has $v+b$ linearly independent rows we will consider the $v+b+2$ maximal independent sets in Lemma 1.

We create the first v rows of A_G by equating the weight of each of the maximal independent sets M_P to the weight of the maximal independent set consisting of all the lines of G. After subtracting, we obtain v equations of the form

$$f(P) - f(\ell_1) - f(\ell_2) - \cdots - f(\ell_r) = 0, \tag{1}$$

where each ℓ_i is incident with P. It follows that, after organizing the columns of A_G by putting point-vertices first and then line-vertices, the “first” v rows of A_G are

$$\begin{bmatrix} I_v & -C \end{bmatrix},$$

where C is the incidence matrix of $\mathrm{Levi}_{(v_r,b_k)}$.

In order to obtain the next b rows of A_G, we will consider maximal independent sets of the form $M_{P,Q}$. For any given line ℓ of (v_r, b_k), we choose (any) two points on it. We will denote these two points as P_1 and P_2. We then consider the maximal independent set M_{P_1,P_2} and equate its weight to the weight of the maximal independent set M_{P_1}. After subtracting, we obtain an equation of the form

$$f(P_2) - f(\ell_1) - f(\ell_2) - \cdots - f(\ell_r) + f(\ell) = 0, \tag{2}$$

where each ℓ_i is incident with P_2.

Note that subtracting (1) (with $P = P_2$) from (2) yields $f(\ell) = 0$. Since ℓ was arbitrary, we get $f(\ell) = 0$ for every line in (v_r, b_k). It follows that since subtracting equations is just a different way to describe row operations in A_G, we get that the "first" $v+b$ rows of A_G (after a few row operations) are

$$\begin{bmatrix} I_v & -C \\ \mathbf{0} & I_b \end{bmatrix}.$$

Note that addition and subtraction were the only two (row) operations needed to obtain the matrix above. Hence, the first $v+b$ rows of A_G do not change depending on the characteristic of F.

Since the determinant of the matrix above is nonzero, the rank of A_G is maximal, and thus $\mathrm{wcdim}(\mathrm{Levi}_{(v_r,b_k)}) = 0$. □

3. Possible generalizations

In this section, we study possible generalizations of Theorem 1. This will be done by providing a few results and by introducing objects to which this theorem could be extended. We begin by proving that Theorem 1 cannot be extended to configurations having exactly two lines being incident with every point. This will be done by an example that considers (v_2)-configurations.

We first notice that a (v_2)-configuration is a disjoint union of polygons/cycles. This is convenient because disjoint unions of graphs behave well with respect to the well-covered dimension. In fact, Lemma 5 in [Brown and Nowakowski 2005] says

$$\mathrm{wcdim}(G \cup H) = \mathrm{wcdim}(G) + \mathrm{wcdim}(H),$$

where $\cup$ stands for disjoint union.

Since we know that $\mathrm{Levi}_{C_n} = C_{2n}$, we get the following lemma.

Lemma 2. *Let $\mathcal{C}$ be a (v_2)-configuration. Then,*

$$\mathcal{C} = \bigcup_{i=1}^{t} C_{n_i},$$

where $n_i > 2$, for all $1 \le i \le t$. Moreover,

$$\mathrm{wcdim}(\mathrm{Levi}_{\mathcal{C}}) = \sum_{i=1}^{t} \mathrm{wcdim}(C_{2n_i}).$$

Finally, we notice that Theorem 5 in [Birnbaum et al. 2014] implies

$$\mathrm{wcdim}(C_{2n}) = \begin{cases} 2 & \text{if } n = 3, \\ 0 & \text{if } n \ge 4. \end{cases}$$

Next is an immediate corollary of that same theorem, together with our Lemma 2.

Corollary 1. *The well-covered dimension of* $\mathrm{Levi}_{\mathcal{C}}$ *is even for all* (v_2)*-configurations* $\mathcal{C}$*. Moreover, for every* $n \in \mathbb{N}$, *there is a* (v_2)*-configuration*, $\mathcal{C}_n$, *such that*

$$\mathrm{wcdim}(\mathrm{Levi}_{\mathcal{C}_n}) = 2n.$$

In particular, *the sequence* $\{\mathrm{wcdim}(\mathrm{Levi}_{\mathcal{C}_n})\}_{n=1}^{\infty}$ *is unbounded.*

We conclude that Theorem 1 cannot be expanded to the case $r = 2$. However, it is still an open problem to find the well-covered dimension of all Levi graphs of (v_2, b_k)-configurations.

Of course, the study of the well-covered dimension of Levi graphs of configurations not of the form (v_r, b_k) is also an interesting open problem.

Block designs are another family of objects that could be studied to attempt a generalization of Theorem 1. These objects can be much less "geometric" than (v_r, b_k)-configurations, given that they are obtained after relaxing items (3) and (4) in Definition 2. In order to be more precise, we provide the following definition.

Definition 5. Let $\lambda, t \geq 1$. A t-(v, k, λ)-design (or t-design), is an incidence structure of points and blocks with the following properties:

(1) There are v points.

(2) Each block is incident with k points.

(3) Any t points are incident with λ common blocks.

It is easy to see that a 1-(v, k, λ)-design is a (v_λ, b_k)-configuration, where $b = v\lambda/k$. Moreover, a 2-$(v, k, 1)$-design is a configuration in which every pair of points are "collinear". For $t > 1$ and $\lambda > 1$, the obvious definition of the Levi graph of a t-design would yield a multigraph. This apparent setback is not so much of a problem since having one edge or multiple edges between two vertices would mean the same thing when looking for maximal independent sets. We claim that the ideas used to prove Theorem 1 can be generalized to be applicable to block designs.

Finally, in this work, we studied the well-covered space of the Levi graph of a (v_r, b_k)-configuration. We propose, as an interesting open problem, the study of configurations via understanding the well-covered spaces of their collinearity graphs (in which points in a configuration are defined as vertices, and adjacency occurs if and only if the points are collinear). The third author is currently working on a particular case of this problem: generalized quadrangles.

Acknowledgments

The authors would like to thank the referee for his/her suggestions. We believe they greatly improved the exposition in this article. We gratefully acknowledge the NSF for their financial support (Grant #DMS-1156273), and the REU program at California State University, Fresno.

References

[Batten 1997] L. M. Batten, *Combinatorics of finite geometries*, 2nd ed., Cambridge University Press, 1997. MR 99c:51001 Zbl 0885.51012

[Birnbaum et al. 2014] I. Birnbaum, M. Kuneli, R. McDonald, K. Urabe, and O. Vega, "The well-covered dimension of products of graphs", *Discuss. Math. Graph Theory* **34**:4 (2014), 811–827. MR 3268692 Zbl 1303.05164

[Bondy and Murty 1976] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*, Elsevier, New York, 1976. MR 54 #117 Zbl 1226.05083

[Brown and Nowakowski 2005] J. I. Brown and R. J. Nowakowski, "Well-covered vector spaces of graphs", *SIAM J. Discrete Math.* **19**:4 (2005), 952–965. MR 2006j:05155 Zbl 1104.05052

[Caro and Yuster 1999] Y. Caro and R. Yuster, "The uniformity space of hypergraphs and its applications", *Discrete Math.* **202**:1-3 (1999), 1–19. MR 2000b:05100 Zbl 0932.05069

[Caro et al. 1998] Y. Caro, M. N. Ellingham, and J. E. Ramey, "Local structure when all maximal independent sets have equal weight", *SIAM J. Discrete Math.* **11**:4 (1998), 644–654. MR 99g:05161 Zbl 0914.05061

[Grünbaum 2009] B. Grünbaum, *Configurations of points and lines*, Graduate Studies in Mathematics **103**, Amer. Math. Soc., Providence, RI, 2009. MR 2011j:52001 Zbl 1205.51003

Received: 2014-10-04 Revised: 2014-12-28 Accepted: 2015-01-02

jessica.hauschild@kwu.edu
Department of Math and Physics, Kansas Wesleyan University,
100 East Claflin Avenue, Salina, KS 91711-5901, United States

jortiz@g.hmc.edu
Department of Mathematics, Harvey Mudd College, 301
Platt Boulevard, Claremont, CA 91711-5901, United States

ovega@csufresno.edu
Department of Mathematics, California State University,
Fresno, Peters Business Building, 5245 North Backer Avenue
M/S PB108, Fresno, CA 93740-8001, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the Involve website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *Involve* are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@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

2015 vol. 8 no. 5