involve

a journal of mathematics

mathematical sciences publishers

2012 vol. 5, no. 1

involve

msp.berkeley.edu/involve

PRODUCTION

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

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

The subscription price for 2012 is US $105/year for the electronic version, and $145/year (+$35 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94704-3840, USA.

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

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

PUBLISHED BY

mathematical sciences publishers

http://msp.org/

A NON-PROFIT CORPORATION

Typeset in LaTeX

INVOLVE 5:1(2012)

Elliptic curves, eta-quotients and hypergeometric functions

David Pathakjee, Zef RosnBrick and Eugene Yoong

(Communicated by Kenneth S. Berenhaut)

The well-known fact that all elliptic curves are modular, proven by Wiles, Taylor, Breuil, Conrad and Diamond, leaves open the question whether there exists a nice representation of the modular form associated to each elliptic curve. Here we provide explicit representations of the modular forms associated to certain Legendre form elliptic curves ${}_2E_1(\lambda)$ as linear combinations of quotients of Dedekind's eta-function. We also give congruences for some of the modular forms' coefficients in terms of Gaussian hypergeometric functions.

1. Introduction and statement of results

Wiles and Taylor [1995] proved that all semistable elliptic curves over $\mathbb{Q}$ are modular. Their result was later extended by Breuil, Conrad, Diamond and Taylor [Breuil et al. 2001] to all elliptic curves over $\mathbb{Q}$.

This correspondence allows facts about elliptic curves to be proven using modular forms, and vice versa. (See [Koblitz 1993] for more background on the theory of elliptic curves and modular forms.)

Let E be an elliptic curve over $\mathbb{Q}$. If $q := e^{2\pi iz}$, GF(p) is the finite field with p elements, and $N(p)$ is the number of points on E over GF(p), then the modularity theorem implies that there exists a corresponding weight-2 newform $f(z) = \sum_{n=1}^{\infty} a(n)q^n$ such that if p is a prime of good reduction, then $a(p) = 1 + p - N(p)$.

For example, if $\eta(z)$ is Dedekind's eta-function,

$$\eta(z) := q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1 - q^n),$$

then the elliptic curves $y^2 = x^3 + 1$ and $y^2 = x^3 - x$ have the corresponding modular forms $\eta(6z)^4$ and $\eta(4z)^2\eta(8z)^2$, respectively; see [Martin and Ono 1997].

MSC2000: primary 11F11, 11F20, 11G05; secondary 11T24, 33C99.

Keywords: number theory, elliptic curves, eta quotients, hypergeometric functions.

The authors wish to thank the NSF for supporting this research, and Ken Ono and Marie Jameson for their invaluable advice.

It is natural to ask which elliptic curves have corresponding modular forms that are quotients of eta-functions. Martin and Ono [1997] have answered this question by listing all such *eta-quotients*

$$f(z) = \prod_{\delta} \eta(\delta z)^{r_\delta} \quad (\delta, r_\delta \in \mathbb{Z})$$

which are weight-2 newforms, and they gave corresponding modular elliptic curves. (For more on the theory of eta-quotients, see [Ono 2004, Section 1.4].)

We show, for certain values of $\lambda \in \mathbb{Q} \setminus \{0, 1\}$, that the elliptic curves ${}_2E_1(\lambda)$ defined by

$$ {}_2E_1(\lambda) : y^2 = x(x-1)(x-\lambda) \tag{1-1}$$

correspond to modular forms which are linear combinations of eta-quotients.

Remark. The proof of Theorem 1.1 will make clear how one can generate many more such examples.

Let

$$f_\lambda(z) := \sum_{n=1}^{\infty} {}_2a_1(n;\lambda)q^n \tag{1-2}$$

be the weight-2 newform corresponding to the elliptic curve ${}_2E_1(\lambda)$. It will be convenient to express eta-quotients using the notation

$$\left[\prod_{\delta} \delta^{r_\delta}\right] := \prod_{\delta} \eta(\delta z)^{r_\delta}. \tag{1-3}$$

For example, in place of $\dfrac{\eta(2z)^2\eta(4z)^2\eta(5z)\eta(40z)}{\eta(z)\eta(8z)}$ we write $[1^{-1}2^2 4^2 5^1 8^{-1} 40^1]$.

Theorem 1.1. *If* $\lambda \in \left\{\frac{27}{16}, 5, \frac{81}{49}, -\frac{7}{25}\right\}$, *then* ${}_2E_1(\lambda)$ *corresponds to the modular forms given here*:

λ	conductor N	eta-quotient $f_\lambda(z)$
$\frac{27}{16}$	33	$[1^2 11^2] + 3\cdot[3^2 33^2] + 3\cdot[1^1 3^1 11^1 33^1]$
5	40	$[1^{-1}2^2 4^2 5^1 8^{-1} 40^1] + [1^1 5^{-1} 8^1 10^2 20^2 40^{-1}]$
$\frac{81}{49}$	42	$2\cdot[1^{-1}2^2 3^1 7^2 14^{-1} 42^1] - 3\cdot[3^1 6^1 21^1 42^1]$ $+ [2^1 3^2 6^{-1} 7^1 21^{-1} 42^2] + [1^1 3^{-1} 6^2 14^1 21^2 42^{-1}]$
$-\frac{7}{25}$	70	$[1^{-1}2^2 5^2 7^{-1} 10^{-1} 14^2 35^2 70^{-1}] - [1^2 2^{-1} 5^{-1} 7^2 10^2 14^{-1} 35^{-1} 70^2]$

We show, for all $\lambda \in \mathbb{Q} \setminus \{0, 1\}$, that the Fourier coefficients of all $f_\lambda(z)$ satisfy an interesting hypergeometric congruence. For a prime p and an integer n, define

$\operatorname{ord}_p(n)$ to be the power of p dividing n, and if $\alpha = \frac{a}{b} \in \mathbb{Q}$, then set $\operatorname{ord}_p(\alpha) = \operatorname{ord}_p(a) - \operatorname{ord}_p(b)$. We show that with this notation, the numbers ${}_2a_1(p;\lambda)$ satisfy the following congruences.

Theorem 1.2. *Let* $\lambda \notin \{0, 1\}$ *be rational and let* $p = 2f + 1$ *be an odd prime such that* $\operatorname{ord}_p(\lambda(\lambda-1)) = 0$. *Then*

$$ {}_2a_1(p;\lambda) \equiv (-1)^{\frac{p+1}{2}}(p-1)\sum_{k=0}^{f}\binom{f+k}{k}\binom{f}{k}(-\lambda)^k \pmod{p}. $$

Remarks. In light of Theorem 1.1, this implies that the congruence in Theorem 1.2 holds for the coefficients of the linear combinations of eta-quotients given above.

- A well-known theorem of Hasse states that for every prime p,

$$ |a(p)| < 2\sqrt{p}. $$

Theorem 1.2 therefore determines ${}_2a_1(p;\lambda)$ uniquely for primes $p > 16$.

Example. Consider $\lambda = \frac{27}{16}$. Then $\lambda(\lambda-1) = \frac{3^3\cdot 11}{2^8}$ and so for $p \notin \{2, 3, 11\}$ prime we observe the congruence by inspecting the coefficients of ${}_2E_1\left(\frac{27}{16}\right)$ for applicable primes $p < 30$, where $B(p;\lambda)$ is defined to be the right-hand side of the congruence in Theorem 1.2:

p	${}_2a_1\left(p;\frac{27}{16}\right)$	$B\left(p;\frac{27}{16}\right)$
5	$-2 \equiv 3 \pmod{5}$	3
7	$4 \equiv 4 \pmod{7}$	4
13	$-2 \equiv 11 \pmod{13}$	11
17	$-2 \equiv 15 \pmod{17}$	15
19	$0 \equiv 0 \pmod{19}$	0
23	$8 \equiv 8 \pmod{23}$	8
29	$-6 \equiv 23 \pmod{29}$	23

2. Elliptic curves and modular forms

In this section we prove Theorem 1.1. If E is an elliptic curve over $\mathbb{Q}$, then its conductor N is a product of the primes p of bad reduction for E, with exponents determined by the extent to which E is singular over $\mathrm{GF}(p)$. (An algorithm by Tate for computing conductors is given in [Cremona 1997].) Moreover, the modularity theorem implies that the modular form $f(z)$ corresponding to E is an element of $S_2(\Gamma_0(N))$. In particular, for an elliptic curve ${}_2E_1(\lambda)$, proving the correctness of any representation of $f_\lambda(z)$ in terms of eta-quotients amounts to checking that the given eta-quotients are elements of $S_2(\Gamma_0(N))$ and checking a finite number of coefficients of their Fourier expansions against those of f_λ.

We first provide a formula for the dimension of the space of cusp forms of weight 2 and level N, $S_2(\Gamma_0(N))$. We then show that the eta-quotients making up the linear combinations are elements of $S_2(\Gamma_0(N))$ and use the dimension formula to show that equality of two elements of $S_2(\Gamma_0(N))$ always depends only on some finite set of coefficients.

The linear combinations of eta-quotients in this paper were generated by the following algorithm:

(1) Given a rational number $\lambda \notin \{0, 1\}$, compute the conductor N of ${}_2E_1(\lambda)$. (The modular form corresponding to ${}_2E_1(\lambda)$ will be an element of $S_2(\Gamma_0(N))$.)

(2) Compute $\dim_{\mathbb{C}} S_2(\Gamma_0(N))$.

(3) Generate eta-quotients which are elements of $S_2(\Gamma_0(N))$.

(4) Attempt to construct a basis for $S_2(\Gamma_0(N))$ using these eta-quotients.

Of course, once one is armed with a basis of eta-quotients for $S_2(\Gamma_0(N))$, it is simple to express $f_\lambda(z)$ in terms of this basis.

Dimension of $S_2(\Gamma_0(N))$. It will be useful to know not only that $S_2(\Gamma_0(N))$ is finite-dimensional for every positive integer N, but also its exact dimension $d_N := \dim_{\mathbb{C}} S_2(\Gamma_0(N))$.

The following formula for d_N is a simplification of [Ono 2004, Theorem 1.34], which gives a formula for the quantity $\dim_{\mathbb{C}} S_k(\Gamma_0(N), \chi) - \dim_{\mathbb{C}} M_{2-k}(\Gamma_0(N), \chi)$, in the case where $k = 2$ and $\chi = \epsilon$ is the trivial character modulo N.

Proposition 2.1. *If N is a fixed positive integer and $r_p := \operatorname{ord}_p(N)$, define*

$$\lambda_p := \begin{cases} p^{\frac{r_p}{2}} + p^{\frac{r_p}{2}-1} & \textit{if } r_p \equiv 0 \pmod 2, \\ 2p^{\frac{r_p-1}{2}} & \textit{if } r_p \equiv 1 \pmod 2. \end{cases}$$

With this notation,

$$d_N = 1 + \frac{N}{12}\prod_{p\mid N}(1+p^{-1}) - \frac{1}{2}\prod_{p\mid N}\lambda_p - \frac{1}{4}\sum_{\substack{x \pmod N \\ x^2+1\equiv 0 \pmod N}} 1 - \frac{1}{3}\sum_{\substack{x \pmod N \\ x^2+x+1\equiv 0 \pmod N}} 1.$$

Proof. This follows from [Ono 2004, Theorem 1.34], noting that the conductor of the trivial character is 1 and that $M_0(\Gamma_0(N), \epsilon)$ is the space of constant functions and hence has dimension 1. □

Proof of Theorem 1.1. Let N be the conductor of $E = {}_2E_1(\lambda)$ and let $d_N = \dim_{\mathbb{C}} S_2(\Gamma_0(N))$ as before. Conditions under which an eta-quotient is an element of $S_2(\Gamma_0(N)))$ are provided in [Ono 2004, Theorems 1.64 and 1.65]: If $f(z) = \prod_{\delta\mid N} \eta(\delta z)^{r_\delta}$ is an eta-quotient which vanishes at each cusp of $\Gamma_0(N)$, such that the pairs (δ, r_δ) satisfy $\sum_{\delta\mid N} r_\delta = 4$, $\sum_{\delta\mid N} \delta r_\delta \equiv 0 \pmod{24}$, and $\sum_{\delta\mid N} \frac{N}{\delta} r_\delta \equiv 0 \pmod{24}$, then $f(z) \in S_2(\Gamma_0(N))$. The order of vanishing of such an $f(z)$ at the cusp $\frac{c}{d}$ is given by [Ono 2004, Theorem 1.65] as

$$\frac{N}{24}\sum_{\delta \mid N} \frac{\gcd(d,\delta)^2 r_\delta}{\gcd(d,\frac{N}{d})d\delta}. \tag{2-1}$$

It is straightforward to check that the formula above gives a positive order of vanishing for each eta-quotient at each cusp, that each eta-quotient satisfies the given congruence conditions, and that the r_δ of each eta-quotient sum to 4. These conditions guarantee that each eta-quotient appearing in the table above lies in $S_2(\Gamma_0(N))$.

The eta-quotients given for $\lambda = \frac{27}{16}$ form a basis for $S_2(\Gamma_0(33))$. Similarly, for $\lambda = 5$, the given eta-quotients along with $[2^2 10^2]$ form a basis; for $\lambda = \frac{81}{49}$ the given eta-quotients along with $[1^{-1}2^2 3^2 6^{-1}7^{-1}14^2 21^2 42^{-1}]$ form a basis; and for $\lambda = -\frac{7}{25}$ a complete basis is

$$\begin{aligned}\{&[5^2 7^2],\ [1^{-1}2^2 7^2 10^1 14^{-1}35^1],\ [10^2 14^2],\\ &[1^2 2^{-1}5^1 7^{-1}14^2 70^1],\ [1^2 2^{-1}5^{-1}7^2 10^2 14^{-1}35^{-1}70^2],\\ &[1^1 5^1 7^1 35^1],\ [1^1 5^2 10^{-1}14^1 35^{-1}70^2],\ [5^1 10^1 35^1 70^1],\ [1^{-1}2^2 5^1 7^1 35^{-1}70^2]\}.\end{aligned}$$

To see this, let $g_{i,j}$ be the j-th Fourier coefficient of the i-th basis vector g_i and define $t_1 < \cdots < t_{d_N}$ to be the first ascending set of indices for which the vectors $\{(g_{i,t_j})_{j=1}^{d_N}\}_{i=1}^{d_N}$ are linearly independent. One can find such a sequence by direct computation of the Fourier coefficients and inspection of the matrices $[g_{i,t_j}]_{i,j=1}^{d_N}$ for various choices of small $t_1 < \cdots < t_{d_N}$.

Now let $v_i = (g_{i,t_1}, \ldots, g_{i,t_{d_N}})$ and let $b_1, \ldots, b_{d_N}$ be a basis for $S_2(\Gamma_0(N))$. If we have $h_1, h_2 \in S_2(\Gamma_0(N))$ with equal t_i-th coefficients, then these coefficients are zero in the difference $h_1 - h_2$. But $h_1 - h_2$ can be written as a linear combination $\sum c_i b_i$ of basis elements, for constants c_i. Hence $\sum c_i v_i = 0$ in $\mathbb{R}^{d_N}$, so by linear independence all $c_i = 0$, and thus $h_1 - h_2 = 0$. It therefore suffices to check that the coefficients of f_λ on $q^{t_1}, \ldots, q^{t_{d_N}}$ match the coefficients that result from the linear combination of eta-quotients. □

Remark. In practice, these computations can be done using a computer algebra system such as SAGE.

Example. We show that the modular form corresponding to ${}_2E_1\bigl(\frac{27}{16}\bigr)$ is

$$g(z) := [1^2 11^2] + 3\cdot[3^2 33^2] + 3\cdot[1^1 3^1 11^1 33^1].$$

For convenience, let $G = \{[1^2 11^2], [3^2 33^2], [1^1 3^1 11^1 33^1]\}$ be the set of eta-quotients making up the linear combination $g(z)$. The conductor of ${}_2E_1\bigl(\frac{27}{16}\bigr)$ is 33 and so the corresponding modular form $f_{\frac{27}{16}}(z)$ is an element of $S_2(\Gamma_0(33))$.

To show that $g(z)$ is also an element of $S_2(\Gamma_0(33))$, it suffices to show that $G \subset S_2(\Gamma_0(33))$. Take $g_i(z) \in G$. By [Ono 2004, Theorem 1.64], $g_i(z)$ is a modular form of weight 2 for $\Gamma_0(33)$. By [Ono 2004, Theorem 1.65], $g_i(z)$ vanishes at all cusps of $\Gamma_0(33)$, and thus $g_i(z) \in S_2(\Gamma_0(33))$.

Since $\mathrm{ord}_3(33) = \mathrm{ord}_{11}(33) = 1$, we have $\lambda_3 = \lambda_{11} = 2$ and evaluation of the dimension formula in Proposition 2.1 gives

$$\begin{aligned}
&\dim_{\mathbb{C}} S_2(\Gamma_0(33))\\
&\quad = 1 + \frac{33}{12}\prod_{p \mid 33}(1+p^{-1}) - \frac{1}{2}\prod_{p\mid 33}\lambda_p - \frac{1}{4}\sum_{\substack{x \pmod{33}\\ x^2+1\equiv 0 \pmod{33}}} 1 - \frac{1}{3}\sum_{\substack{x \pmod{33}\\ x^2+x+1\equiv 0 \pmod{33}}} 1\\
&\quad = 1 + \tfrac{33}{12}\left(1+\tfrac{1}{3}\right)\left(1+\tfrac{1}{11}\right) - \tfrac{1}{2}(\lambda_3)(\lambda_{11}) - \tfrac{1}{4}(0) - \tfrac{1}{3}(0)\\
&\quad = 3.
\end{aligned}$$

It remains to show that G is a basis for $S_2(\Gamma_0(33))$. Any dependence relation satisfied by the elements of G would imply a dependence relation among their coefficients. It thus suffices to find a set of indices $t_1 < t_2 < t_3$ such that the 3×3 matrix formed by the t_i-th coefficients of these eta-quotients is nonsingular. For this particular λ, the first three coefficients suffice.

This implies that any two elements of $S_2(\Gamma_0(33))$ which agree on the first three coefficients are equal. In fact, we observe that the first three coefficients of the modular form corresponding to ${}_2E_1\left(\frac{27}{16}\right)$ are the same as the first three coefficients of $g(z)$. That is, the coefficients of $g(z) = q + q^2 - q^3 - q^4 + \cdots$ agree with the coefficients of $f_{\frac{27}{16}}(z)$.

3. Gaussian hypergeometric functions and proof of Theorem 1.2

We recall some facts about Gaussian hypergeometric functions over finite fields of prime order and use the Gaussian hypergeometric function ${}_2F_1\left(\begin{smallmatrix}\phi, & \phi\\ & \epsilon\end{smallmatrix} \mid \lambda\right)$ to prove Theorem 1.2.

Gaussian hypergeometric functions. Greene [1987] defined *Gaussian hypergeometric functions* over arbitrary finite fields and showed that they have properties analogous to those of classical hypergeometric functions. We recall some definitions and notation from [Ono 1998] in the case of fields of prime order.

Definition 3.1. If p is an odd prime, $\mathrm{GF}(p)$ is the field with p elements, and A and B are characters of $\mathrm{GF}(p)$, define

$$\binom{A}{B} := \frac{B(-1)}{p}J(A,\bar{B}) = \frac{B(-1)}{p}\sum_{x\in\mathrm{GF}(p)} A(x)\bar{B}(1-x).$$

Furthermore, if $A_0, \dots, A_n$ and $B_1, \dots, B_n$ are characters of $\mathrm{GF}(p)$, define the Gaussian hypergeometric series ${}_{n+1}F_n\left(\begin{smallmatrix} A_0, & A_1, & \dots, & A_n \\ & B_1, & \dots, & B_n \end{smallmatrix} \mid x\right)$ by the following sum over all characters χ of $\mathrm{GF}(p)$:

$${}_{n+1}F_n\left(\begin{smallmatrix} A_0, & A_1, & \dots, & A_n \\ & B_1, & \dots, & B_n \end{smallmatrix} \mid x\right) := \frac{p}{p-1}\sum_{\chi}\binom{A_0\chi}{\chi}\binom{A_1\chi}{B_1\chi}\cdots\binom{A_n\chi}{B_n\chi}\chi(x)$$

In particular, we are concerned with the Gaussian hypergeometric series ${}_2F_1(\lambda)$ defined by

$${}_2F_1(\lambda) := {}_2F_1\left(\begin{smallmatrix} \phi, & \phi \\ & \epsilon \end{smallmatrix} \mid \lambda\right) = \frac{p}{p-1}\sum_{\chi}\binom{\phi\chi}{\chi}^2\chi(\lambda)$$

where ϕ is the quadratic character of $\mathrm{GF}(p)$. It is shown in [Ono 1998] that if $\lambda \in \mathbb{Q} \setminus \{0, 1\}$, then

$${}_2F_1(\lambda) = -\frac{\phi(-1)\,{}_2a_1(p;\lambda)}{p} \tag{3-1}$$

for every odd prime p such that $\mathrm{ord}_p(\lambda(\lambda-1)) = 0$.

In addition, define the generalized Apéry number $D(n; m, l, r)$ for every $r \in \mathbb{Q}$ and every pair of nonnegative integers m and l by

$$D(n; m, l, r) := \sum_{k=0}^{n}\binom{n+k}{k}^m\binom{n}{k}^l r^{lk}.$$

Ono also shows (ibid.) that if $p = 2f+1$ is an odd prime and $w = l+m$, then

$$D(f; m, l, r) \equiv \left(\frac{p}{p-1}\right)^{w-1} {}_wF_{w-1}\left(\begin{smallmatrix} \phi, & \phi, & \dots, & \phi \\ & \epsilon, & \dots, & \epsilon \end{smallmatrix} \mid (-r)^l\right) \pmod{p}. \tag{3-2}$$

Proof of Theorem 1.2. By (3-1) and the fact that $\phi(-1) = (-1)^{\frac{p-1}{2}}$, we have that

$$\frac{p}{p-1}{}_2F_1(\lambda) = \frac{(-1)^{\frac{p+1}{2}}\,{}_2a_1(p;\lambda)}{p-1}.$$

By (3-2), letting $l = m = 1$ (and thus $w = 2$) and $r = -\lambda$, we have

$$\frac{p}{p-1}{}_2F_1(\lambda) \equiv D\left(f; 1, 1, -\lambda\right) \pmod{p}.$$

Combining these two equations and rearranging, we get

$${}_2a_1(p;\lambda) \equiv (-1)^{\frac{p+1}{2}}(p-1)D\left(f; 1, 1, -\lambda\right) \pmod{p}.$$

Since

$$D\left(f; 1, 1, -\lambda\right) = \sum_{k=0}^{n}\binom{f+k}{k}\binom{f}{k}(-\lambda)^k,$$

we have

$$ {}_2a_1(p;\lambda) \equiv (-1)^{\frac{p+1}{2}}(p-1)\sum_{k=0}^{f}\binom{f+k}{k}\binom{f}{k}(-\lambda)^k \pmod{p}. \qquad \square $$

Remark. The binomial product $\binom{f+k}{k}\binom{f}{k}$ can be combined into the multinomial coefficient $\binom{f+k}{k,\,k,\,f-k}$ and so the congruence in Theorem 1.2 can also be written as

$$ {}_2a_1(p;\lambda) \equiv (-1)^{\frac{p+1}{2}}(p-1)\sum_{k=0}^{f}\binom{f+k}{k,k,f-k}(-\lambda)^k \pmod{p}. $$

References

[Breuil et al. 2001] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, "On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises", *J. Amer. Math. Soc.* **14**:4 (2001), 843–939. MR 2002d:11058 Zbl 0982.11033

[Cremona 1997] J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, Cambridge, 1997. MR 99e:11068 Zbl 0872.14041

[Greene 1987] J. Greene, "Hypergeometric functions over finite fields", *Trans. Amer. Math. Soc.* **301**:1 (1987), 77–101. MR 88e:11122 Zbl 0629.12017

[Koblitz 1993] N. Koblitz, *Introduction to elliptic curves and modular forms*, 2nd ed., Graduate Texts in Mathematics **97**, Springer, New York, 1993. MR 94a:11078 Zbl 0804.11039

[Martin and Ono 1997] Y. Martin and K. Ono, "Eta-quotients and elliptic curves", *Proc. Amer. Math. Soc.* **125**:11 (1997), 3169–3176. MR 97m:11057 Zbl 0894.11020

[Ono 1998] K. Ono, "Values of Gaussian hypergeometric series", *Trans. Amer. Math. Soc.* **350**:3 (1998), 1205–1223. MR 98e:11141 Zbl 0910.11054

[Ono 2004] K. Ono, *The web of modularity: arithmetic of the coefficients of modular forms and q-series*, CBMS Regional Conference Series in Mathematics **102**, American Mathematical Society, Providence, RI, 2004. MR 2005c:11053 Zbl 1119.11026

[Taylor and Wiles 1995] R. Taylor and A. Wiles, "Ring-theoretic properties of certain Hecke algebras", *Ann. of Math.* (2) **141**:3 (1995), 553–572. MR 96d:11072 Zbl 0823.11030

Received: 2010-05-03 Revised: 2011-04-22 Accepted: 2011-09-14

pathakjee@wisc.edu
Department of Mathematics,
University of Wisconin - Madison, 480 Lincoln Drive,
Madison, WI 53706-1388, United States

rosnbrick@wisc.edu
Department of Mathematics,
University of Wisconsin - Madison, 480 Lincoln Drive,
Madison, WI 53706-1388, United States

eyoong@uwaterloo.ca
Department of Pure Mathematics, University of Waterloo,
200 University Avenue West, Waterloo, ON, N2L 3G1,
Canada

Trapping light rays aperiodically with mirrors

Zachary Mitchell, Gregory Simon and Xueying Zhao

(Communicated by Joseph O'Rourke)

We construct a configuration of disjoint segment mirrors in the plane that traps a single light ray aperiodically, providing a negative solution to a conjecture of O'Rourke and Petrovici. We expand this to show that any finite number of rays from a source can be trapped aperiodically.

1. Background and statement of results

We consider a point source of light together with a finite collection of disjoint, double-sided segment mirrors in the plane. Light rays travel from the source in fixed directions until they contact a mirror, at which point they reflect according to the laws of geometric optics: the angle of incidence equal to the angle of reflection. Rays that contact mirror endpoints are assumed to die there, and such rays are called *degenerate*. A ray is said to be *trapped* if it never escapes from the convex hull of the mirrors. A ray is said to be trapped *aperiodically* if it reflects at infinitely many distinct mirror points. Such a ray reflects forever but never retraces its own path.

O'Rourke and Petrovici [2001] asked whether the light rays emanating in every direction from a point source could be simultaneously trapped in such a system. The authors show that the set of rays from a source trapped periodically is countable; this can also be shown for the set of degenerate rays. The remaining rays either escape or are trapped aperiodically. Thus, if all the light rays from a point source are trapped by the mirrors, then there must be uncountably many aperiodically trapped rays. The authors' conjecture that mirrors cannot trap a light ray aperiodically would have implied that no mirror system could trap all light from a point source. We provide a counterexample to this conjecture:

Theorem 1. *There is a mirror configuration in the plane that traps a light ray aperiodically.*

David Milovich tells us this was proved independently in 2002 by Ben Stephens, then a graduate student at MIT, but left unpublished; see [O'Rourke 2005].

MSC2000: 37D50, 78A05.

Keywords: trapping light, mirrors, billiards, dynamical systems.

Funded by NSF REU Award #0502205.

We then build upon this construction inductively to show:

Theorem 2. *For any $n \geq 1$, there is a mirror configuration in the plane that traps n distinct aperiodic rays from a single source.*

2. Proofs

The bases for both arguments lie in a fundamental result from mathematical billiards: any billiard path in the square with irrational slope traces out an aperiodic trajectory [Tabachnikov 2005, pp. 25–26]. Our method is to recreate the dynamics of the billiard path in the square using disjoint segment mirrors.

Proof of Theorem 1. The construction will focus around the 2×2 square in the (x, y)-plane defined by $\max\{|x|, |y|\} = 1$, i.e., the square with the four vertices $(\pm 1, \pm 1)$. This will be referred to simply as *the square*. We begin with two segment mirrors, the top and bottom of the square. Fix a point p and an angle θ, measured counterclockwise from horizontal, so that the initial position and direction (p, θ) would define an aperiodic billiard trajectory in the square. Although this trajectory is aperiodic, it consists of only four distinct directions: θ, $-\theta$, $\pi - \theta$, and $\pi + \theta$.

We place six additional mirrors outside the opened square, four horizontal and two vertical (Figure 1), in such a way that a ray that exits the opened square in one of these four directions necessarily returns to the same point of the square where it exited. Such a ray always hits three mirrors (horizontal, vertical, then horizontal) before returning; thus its direction changes as if it reflects only off of a vertical mirror. Ignoring the path outside the square, the trajectory behaves as though it has reflected off the square's (absent) vertical edge (Figure 2).

A set of mirrors with this property is given explicitly at the end of this proof, as can be verified by simple trigonometry. Within the square, the light ray (p, θ) travels as though all four mirrors of the square were in place. In particular, this ray is trapped aperiodically.

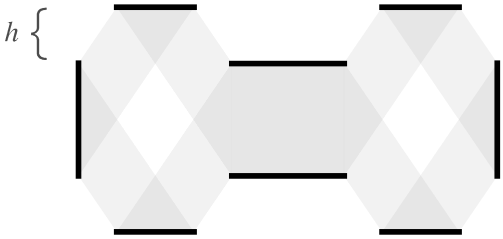


Figure 1. The mirror configuration, along with four strips of parallel rays, representing all four possible directions of escape, shown exiting the central square and returning.

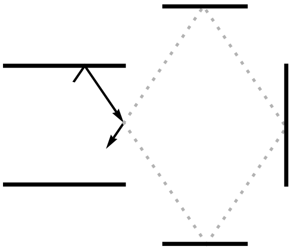


Figure 2. The ray appearing to reflect off the right vertical edge of the square.

We conclude with the mirror coordinates. Fix $h > 0$; this parameter gives the height difference between the top of the square and the height of two higher horizontal mirrors. "$P \sim Q$" will denote the closed segment mirror from P to Q. Our configuration is symmetric about both the x- and y-axes. Coordinates for the three mirrors that meet the upper right quadrant of the plane are given: (the remaining 5 can be obtained via symmetry)

$$\begin{gathered}(-1, 1) \sim (1, 1),\\ (1 + h\,|\cot\theta|, 1 + h) \sim (1 + (h + 2)|\cot\theta|, 1 + h),\\ (1 + (2h + 2)|\cot\theta|, -1) \sim (1 + (2h + 2)|\cot\theta|, 1).\end{gathered} \qquad \square$$

Proof of Theorem 2. Suppose $0 < \theta_1 < \pi/2$ and (p, θ_1) is trapped as described in Theorem 1 in a configuration with $h = h_1$. To trap an additional ray from the source, we choose another aperiodic direction θ_2 and height h_2 carefully to ensure that the new construction does not interact with the old.

Intuitively, we choose θ_2 to be very steep, thus when the ray (p, θ_2) escapes from the central square, it will also escape through the gaps of the initial construction. So for θ_2, we require that the initial position and direction (p, θ_2) follows an aperiodic trajectory in the square, that $\theta_1 < \theta_2 < \pi/2$, and that the ray starting at the bottom right corner of the square and traveling in the direction θ_2 escapes from the original mirror system without reflections. By symmetry, this guarantees that *any* ray that exits the opened square in one of the four possible directions θ_2, $-\theta_2$, $\pi - \theta_2$, or $\pi + \theta_2$ will not reflect off any of the original mirrors. By choosing the height h_2 for the horizontal mirrors to be sufficiently large, we can ensure that after the ray (p, θ_2) exits the central square, its path will completely surround the original mirrors; formally, this is achieved when $h_2 > (h_1{+}1)\cot\theta_1 \tan\theta_2$, as can be verified with straightforward trigonometry.

This will guarantee that the new ray will not hit the original mirrors and that the new mirrors will not interfere with the original ray. (Compare Figure 1 with

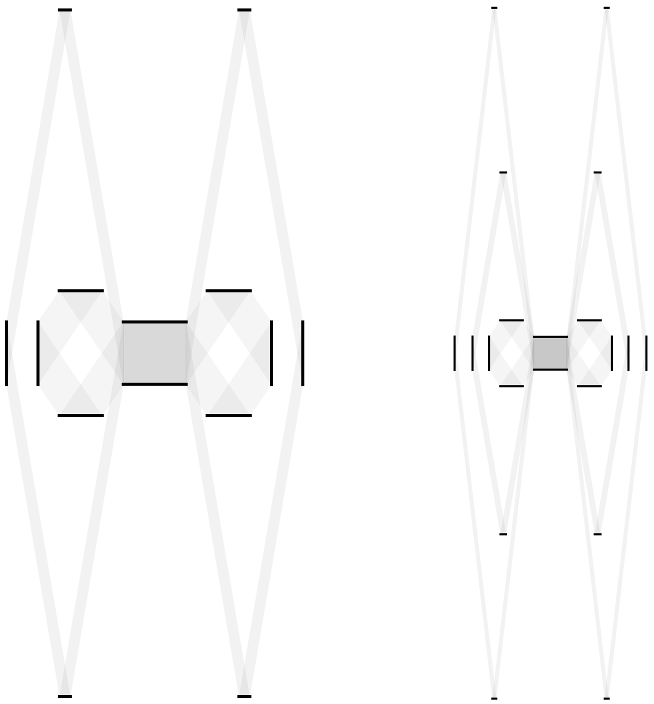


Figure 3. Examples for the cases $n = 2$ (left) and $n = 3$ (right).

the left part of Figure 3). In this way, the two rays are simultaneously trapped aperiodically.

This process can be continued. If θ_i and h_i (for $1 \leq i < n$) have been chosen in this way, then choose θ_n such that (p, θ_n) follows an aperiodic billiard path in the square, that $\theta_{n-1} < \theta_n < \pi/2$, and that the ray from $(1, -1)$ in the direction θ_n escapes from the system without reflection. Choose h_n large enough to ensure that the path of the new ray will completely surround the original system — again, this is accomplished when $h_n > (h_{n-1} + 1) \cot \theta_{n-1} \tan \theta_n$. The ray (p, θ_n) is now trapped aperiodically, as are the previous $n - 1$ rays. Inductively, we can trap any finite number of rays aperiodically. □

3. Further remarks

It may be of interest to strengthen the second theorem. In its original form, we had to choose particular rays which avoided mirrors already in place. This allowed for an easy inductive proof but can be avoided by a direct approach. The finite collection of directions (to be trapped aperiodically) can be arbitrary:

Any finite collection of rays from a source can be trapped aperiodically with finitely many disjoint segment mirrors.

We omit a formal proof but offer an outline. By rotating the plane about the point source, we show that we may assume that each ray's direction is irrational — that is, that they have irrational slope. Suppose $\{\theta_i : 1 \leq i \leq n\}$ is a finite collection of angles, which we can identify with rays eminating from the source. For $1 \leq i \leq n$, let R_i denote the set of rotations that rotate the plane in such a way that the angle θ_i becomes a rational direction after rotation. The set of rational directions is countable, and given a rational direction ρ, there is a unique rotation sending θ_i to ρ. So R_i is in bijection with the set of rational directions, hence R_i is countable. Thus, the set of all rotations that send *any* θ_i to a rational direction is countable, since it is the finite union $\bigcup_{i=1}^{n} R_i$. Because there are uncountably many rotations, some rotation does not send any θ_i to a rational direction — so this rotation sends each θ_i to an irrational direction. After applying such a rotation, we may assume that all of the given rays have irrational slopes. Hence (after rotating in such a way) each follows an aperiodic billiard path in the square.

We then proceed as in Theorem 2, creating n copies of the construction of Theorem 1. In this case, however, the angles θ_i are predetermined, so only the heights h_i can be adjusted. This degree of freedom is enough. Intuitively, as h_i increases, the mirror configuration expands. Provided each h_i is sufficiently large and they differ from one another by a sufficiently large amount, the mirror configurations will not interfere with one another (as in Figure 3). In such a system, each ray is aperiodically trapped.

These constructions do not answer the larger questions of trapping light (namely, if all light from a source can be trapped), but they do bring to the forefront some additional lines of inquiry. We've shown that the cardinality of aperiodically trapped rays can be any finite number, but must this cardinality be finite? Or, a weaker statement, must this cardinality be countable? These questions were originally posed in [O'Rourke and Petrovici 2001]. A positive answer to either would resolve the larger question of whether all light from a point source can be trapped with segment mirrors.

We would also like to mention that the approach used to prove Theorem 1 can be applied to polygons other than the square; our original proof was based on a quadrilateral with a nonperiodic billiard path constructed in [Galperin 1983].

Acknowledgements

This work is a result of the Mathematics REU at Pennsylvania State University, 2009. The authors would like to thank Joseph O'Rourke and David Milovich for their helpful correspondence as well as Misha Guysinsky, Andrey Gogolev, and others at Penn State for their support.

References

[Galperin 1983] G. A. Galperin, “Nonperiodic and not everywhere dense billiard trajectories in convex polygons and polyhedrons”, *Comm. Math. Phys.* **91**:2 (1983), 187–211. MR 85a:58082

[O’Rourke 2005] J. O’Rourke, “The Open Problem Project, Problem 31: trapping light rays with segment mirrors”, Oct 2005, available at http://maven.smith.edu/~orourke/TOPP/P31.html.

[O’Rourke and Petrovici 2001] J. O’Rourke and O. Petrovici, “Narrowing light rays with mirrors”, pp. 137–140 in *Proc. 13th Canad. Conf. Comput. Geom.* (Waterloo, Ont., 2001), 2001.

[Tabachnikov 2005] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library **30**, American Mathematical Society, Providence, RI, 2005. MR 2006h:51001 Zbl 1119.37001

Received: 2010-05-04 Revised: 2011-06-25 Accepted: 2011-07-06

zachary.mitchell@hope.edu *Department of Mathematics, Hope College,*
Holland, MI 49423, United States

gregory.g.simon@gmail.com *Department of Mathematics, University of California,*
Santa Cruz, Santa Cruz, CA 95064, United States

zhao23x@mtholyoke.edu *Department of Mathematics, Mount Holyoke College,*
South Hadley, MA 01075, United States

A generalization of modular forms

Adam Haque

(Communicated by Ken Ono)

We prove a transformation equation satisfied by a set of holomorphic functions with rational Fourier coefficients of cardinality $2^{\aleph_0}$ arising from modular forms. This generalizes the classical transformation property satisfied by modular forms with rational coefficients, which only applies to a set of cardinality $\aleph_0$ for a given weight.

Modular forms play a crucial role in number theory, complex analysis, and geometry. However, from a set-theoretic point of view, the $\mathbb{Q}$-vector space $M_r(\Gamma)$ of holomorphic modular forms of a given weight r on $\Gamma = \mathrm{SL}(2, \mathbb{Z})$ is only a small subset of the meromorphic functions of $q = e^{2\pi i z}$ on the open unit disc D centered at the origin of the complex plane with rational power series coefficients. This is because the set of all modular forms of a given weight r with rational Fourier coefficients is countable (has cardinality $\aleph_0$), as can be seen from the fact that the algebra of all modular forms on Γ over $\mathbb{Q}$ is finitely generated by modular forms with rational coefficients [Ono 2004]. In contrast, since every meromorphic function of $q = e^{2\pi i z}$ on the unit disc D with a pole having at most finite order at $q = 0$ can be represented as a power series of the form

$$g(z) = \sum_{n=-m}^{\infty} a(n) e^{2\pi i n z} \tag{1}$$

uniformly convergent on compact subsets of D and conversely, it is clear that the cardinality of the set of meromorphic or holomorphic functions of $q = e^{2\pi i z}$ with rational power series coefficients is $2^{\aleph_0}$. We discuss this in more detail in the proof of Corollary 4 and Proposition 5.

Since modular forms are only a small subset of the set of all meromorphic functions, it is interesting to ask whether or not it is possible to generalize the definition of modularity so as to encompass a set of functions with cardinality $2^{\aleph_0}$, while still

MSC2000: 11F11, 11F30.

Keywords: generalized modular forms, Dirichlet multiplication, cardinality.

preserving some of the remarkable transformation properties of modular forms. This can be done by allowing the level of the modular form to become infinite.

To be specific, we consider sequences of elements of $\mathrm{SL}(2,\mathbb{Z})$, that is, integer matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ such that $|ad-bc|=1$, where the entries depend on a positive integer k, which will be suppressed from the notation. We will assume that c is an increasing (and therefore unbounded) function of k, and that the quotient d/c approaches a finite limit as $k\to\infty$. Note that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ belongs to the modular group $\Gamma_0(c)$ — by definition, $\Gamma_0(N)$ consists of the matrices in $\mathrm{SL}(2,\mathbb{Z})$ whose lower left entry is a multiple of N. We let $\mathrm{SL}(2,\mathbb{Z})$ act on the upper half-plane $\{z\in\mathbb{C} : \operatorname{Im} z>0\}$ in the usual way:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}.$$

Let r be a positive integer. Let g be a meromorphic function on the upper half-plane with a pole of at most finite order at $z=i\infty$. Suppose there is a sequence $c=c(k)$ with the property that, for *any* sequence $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of the form above *consistent with this choice of* c, the function g satisfies the transformation equation

$$\left(z+\lim_{k\to\infty}\frac{d}{c}\right)^r g(z) = \lim_{k\to\infty} c^{-r} g\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right). \tag{2}$$

In that case we say that g is a *generalized modular form of weight* r, or a *modular form of weight* r *and level infinity*.

To see that this notion is a generalization of traditional modular forms, consider a modular form g of weight r and level N, and take for c the sequence given by $c(k)=Nk$. Any element of any sequence $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ consistent with this choice of c is an element of $\Gamma_0(N)$; therefore, by the definition of a modular form, g satisfies

$$(cz+d)^r g(z) = g\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)$$

for all k. Dividing both sides by c^r and taking the limit as $k\to\infty$ we see that (2) is satisfied.

We will now see how to create uncountably many generalized modular forms with rational coefficients. We recall the definition of Dirichlet multiplication for two sequences $\{h(n)\}$ and $\{C(n)\}$:

$$(h*C)_n = \sum_{d|n} h(d) C\left(\frac{n}{d}\right)$$

We will assume $C(1)\neq 0$ in order to guarantee the existence of the Dirichlet inverse $\{C^{-1}(n)\}$, the inverse of the sequence $\{C(n)\}$ under the operation of Dirichlet multiplication. For efficient notation, we use $\{A_n\}$ and $\{A(n)\}$ interchangeably for any sequence $\{A_n\}$. Here is our main result.

Theorem 1. *Let*

$$\sum_{n=1}^{\infty} C(n)e^{2\pi inz}$$

be a cusp form of even weight $r > 0$ on Γ with $C(1) \neq 0$ and $\{|h(n)|\} \in \ell^1$ (i.e., $\sum_{n=1}^{\infty} h(n)$ is absolutely convergent). Then any holomorphic function on the upper half-plane of the form

$$g(z) = \sum_{n=1}^{\infty} (h * C)_n e^{2\pi inz} = \sum_{n=1}^{\infty} \sum_{d|n} h(d)C\Big(\frac{n}{d}\Big)e^{2\pi inz} \tag{3}$$

is a holomorphic generalized modular form of weight r and level infinity that satisfies the transformation equation

$$\Big(z + \lim_{c(k)\to\infty} \frac{d}{c}\Big)^r g(z) = \lim_{c(k)\to\infty} c^{-r} g\Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\Big); \tag{4}$$

here $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(c)$ with $c(k) = \operatorname{lcm}(1, 2, 3, \ldots, k)$. Thus $g(z)$ satisfies an approximate modular transformation equation, with its accuracy increasing as $c(k) \to \infty$. Here we define (4) *to be such an approximate modular transformation equation.*

This theorem generalizes the result

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^r f(z)$$

when $h(n)$ in (3) is the identity element of Dirichlet multiplication $I(n)$, since in this case $g(z)$ is a cusp form by definition:

$$g(z) = \sum_{n=1}^{\infty} (I * C)_n e^{2\pi inz} = \sum_{n=1}^{\infty} C(n)e^{2\pi inz},$$

$$\Big(z + \frac{d}{c}\Big)^r g(z) = c^{-r} g\Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\Big).$$

Of course, in this case $\{|I(n)|\} \in \ell^1$ since $I(n) = 0$ for $n > 1$, and thus the hypotheses of Theorem 1 are satisfied. We also note that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)z$ approaches the real line as $c(k) \to \infty$, since $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(c)$ implies

$$\frac{az+b}{cz+d} = \frac{a}{c} - \frac{1}{c(cz+d)}, \qquad \lim_{c(k)\to\infty} \operatorname{Im} \frac{az+b}{cz+d} = 0.$$

Proof. We prove Theorem 1 using series of modular forms. In particular we use the cusp form of weight r on Γ given by

$$f(z) = \sum_{n=1}^{\infty} C(n)e^{2\pi inz},$$

where $\{C(n)\}$ is any cusp form coefficient sequence. It is well known that there exist functions that are analytic in the upper half-plane and satisfy the functional equation

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^r f(z),$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and $ad - bc = 1$ (Γ being the modular group). From this property, it is easy to see that if n divides c, that is, if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(n)$, then for positive integer n we have

$$f\Big(n\,\frac{az+b}{cz+d}\Big) = (cz+d)^r f(nz).$$

The Fourier expansion for $f(mz)$

$$f(mz) = \sum_{n=1}^{\infty} C(n) e^{2\pi imnz}$$

is absolutely convergent in the upper half-plane, since $C(n) = O(n^{r/2})$ by a standard argument of Hecke [Apostol 1990]. Assuming $A_m = O(m^p)$ for some natural number p we note that the double series

$$\sum_{m=1}^{\infty} A_m f(mz) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_m C_n e^{2\pi imnz}$$

is absolutely convergent, since both sequences A_m and C_n are bounded by polynomials, while of course $e^{2\pi imnz}$ decays exponentially in absolute value as m or n increases. Hence rearrangement is justified and we can write

$$\sum_{m=1}^{\infty} A_m f(mz) = \sum_{n=1}^{\infty}\sum_{d|n} A(d) C\Big(\frac{n}{d}\Big) e^{2\pi inz} = \sum_{n=1}^{\infty} (A * C)_n e^{2\pi inz}.$$

We also need the identity

$$e^{2\pi iz} = \sum_{m=1}^{\infty} C^{-1}(m) f(mz) \tag{5}$$

where $C^{-1}(m)$ is the Dirichlet inverse of the cusp form coefficients. Assuming absolute convergence, identity (5) follows easily from the following rearrangement:

$$\begin{aligned}\sum_{m=1}^{\infty} C^{-1}(m) f(mz) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} C^{-1}(m) C(n) e^{2\pi imnz} \\ &= \sum_{n=1}^{\infty} (C^{-1} * C)_n e^{2\pi inz} = e^{2\pi iz}.\end{aligned}$$

To prove absolute convergence it is sufficient to prove that $C^{-1}(m)$ is bounded by a polynomial in m. This follows from the fact that $C(n) = O(n^{r/2})$ [Apostol 1990], together with the following lemma:

Lemma 2. *If a sequence $\{C(n)\} \subseteq \mathbb{C}$ with $C(1) \neq 0$ is bounded by a polynomial in n, then its Dirichlet inverse $C^{-1}(m)$ is also bounded by a polynomial in n. In symbols, if $|C(n)| = O(n^{d_1})$ for some $d_1 \in \mathbb{R}$, there exists $d_2 \in \mathbb{R}$ such that $|C^{-1}(n)| = O(n^{d_2})$.*

Proof. We prove this by induction. If $|C(n)| = O(n^{d_1})$, then letting $|C(1)| = P$, we find that there exists $k \in \mathbb{R}$ such that $|C(n)| \leq Pn^k$ for any positive integer n. We use the standard recursive definition

$$C^{-1}(n) = -\frac{1}{C(1)} \sum_{\substack{d|n \\ d<n}} C\Big(\frac{n}{d}\Big) C^{-1}(d), \tag{6}$$

which is equivalent to $(C * C^{-1})_n = I(n)$, where $I(n)$ is the identity element of Dirichlet multiplication. We find that $|C(1)| = P$ implies $|C^{-1}(1)| = 1/P$. We make the inductive hypothesis

$$|C^{-1}(d)| \leq \frac{1}{P} d^{k+2} \quad \text{for all } d < n,\ d \in \mathbb{N}.$$

Using the recursive definition (6) we obtain

$$|C^{-1}(n)| \leq \Big|\frac{1}{C(1)}\Big| \sum_{\substack{d|n \\ d<n}} \Big|C\Big(\frac{n}{d}\Big)\Big| |C^{-1}(d)| \leq \Big|\frac{1}{C(1)}\Big| \sum_{\substack{d|n \\ d<n}} \Big(\frac{n}{d}\Big)^k d^{k+2} \leq \frac{1}{P} n^k \sum_{\substack{d|n \\ d<n}} d^2.$$

So,

$$|C^{-1}(n)| \leq \frac{1}{P} n^k \sum_{\substack{d|n \\ d<n}} d^2 = \frac{1}{P} n^{k+2} \sum_{\substack{d|n \\ d>1}} \frac{1}{d^2} \leq \frac{1}{P} n^{k+2} (\zeta(2) - 1) \leq \frac{1}{P} n^{k+2},$$

where $\zeta(s)$ is the Riemann zeta function. It follows that

$$|C^{-1}(n)| \leq \frac{1}{P} n^{k+2},$$

and this completes the induction. □

Any complex analytic function $J(q)$ can be written as a power series for q in the open unit disc D centered at $q = 0$:

$$J(q) = \sum_{n=0}^{\infty} A_n q^n.$$

Making the substitution $q = e^{2\pi i z}$ with $J(e^{2\pi i z}) = g(z)$ and assuming $J(0) = 0$

for convenience, we find

$$g(z)=\sum_{n=1}^{\infty}A_n e^{2\pi i n z}.$$

Using the absolute convergence of

$$e^{2\pi i n z}=\sum_{m=1}^{\infty}C^{-1}(m)f(mnz)$$

in the upper half-plane, and assuming A_n is bounded by a polynomial in n, we use rearrangement of series to write

$$\begin{aligned}\sum_{n=1}^{\infty}(A*C^{-1})_n f(nz)&=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(A*C^{-1})_n C_m e^{2\pi i m n z}\\&=\sum_{n=1}^{\infty}(A*C^{-1}*C)_n e^{2\pi i n z}=\sum_{n=1}^{\infty}A_n e^{2\pi i n z}=g(z).\end{aligned}$$

This is justified by the discussion above and Lemma 2, which imply that all the series above are absolutely convergent. Now consider the partial sums of the series

$$g_k(z)=\sum_{n=1}^{k}(A*C^{-1})_n f(nz).$$

From this definition, assuming $z=x+iy$, we have

$$|g(z)-g_k(z)|=O(e^{-2\pi k y}). \tag{7}$$

This is because the cusp forms $f(nz)$ decay exponentially as n increases [Shimura 2007], so there exists $M\in\mathbb{R}^+$ such that $|f(nz)|<Me^{-2\pi n y}$ for all n. Hence, as $k\to\infty$ we have by the triangle inequality:

$$\begin{aligned}|g(z)-g_k(z)|&=\left|\sum_{n=k+1}^{\infty}(A*C^{-1})_n f(nz)\right|\\&<Me^{-2\pi k y}\sum_{n=1}^{\infty}(A*C^{-1})_n e^{-2\pi n y}=O(e^{-2\pi k y}).\end{aligned}$$

From the functional equation $f\Big(n\,\dfrac{az+b}{cz+d}\Big)=(cz+d)^r f(nz)$, valid if $n|c$ and $ad-bc=1$, we obtain

$$\Big(z+\frac{d}{c}\Big)^r g_k(z)=c^{-r}\sum_{n=1}^{k}(A*C^{-1})_n f\Big(n\,\frac{az+b}{cz+d}\Big),\quad ad-bc=1,$$

by choosing $c(k)=\mathrm{lcm}[1,2,3,\dots,k]$.

Given this c, we can always choose a, b, d such that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(n)$ for all $n \leq k$ and with d/c approaching a finite limit as $k \to \infty$. For example, one can take $\left(\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right)$ or, more generally,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c\nu+1 & -c\nu^2 \\ c & -c\nu+1 \end{pmatrix}$$

for some integer ν. Hence, we can write

$$\left(z+\frac{d}{c}\right)^r g_k(z) = c^{-r} \sum_{n=1}^{k} (A * C^{-1})_n f\left(n \begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right). \tag{8}$$

This approach, however, does not work for arbitrary holomorphic functions $f(z)$ since the error term

$$c^{-r} \sum_{n=k+1}^{\infty} (A * C^{-1})_n f\left(n \begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)$$

diverges as k and c approach ∞. One way to circumvent this difficulty is to choose an sequence of real numbers $h(n)$ with $\{|h(n)|\} \in \ell^1$, and set

$$A_n = (h * C)_n, \quad \text{or, equivalently,} \quad (A * C^{-1})_n = h(n), \tag{9}$$

so that

$$g(z) = \sum_{n=1}^{\infty} (h * C)_n e^{2\pi i n z}. \tag{10}$$

In this case, A_n is bounded by a polynomial in n and the error term is

$$c^{-r} \sum_{n=k+1}^{\infty} h(n) f\left(n \begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right).$$

Lemma 3. *Let $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be a sequence as on page 16. As $c \to \infty$, we have*

$$\left| f\left(n \begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) \right| < M n^{-r/2} \frac{|cz+d|^r}{(\operatorname{Im} z)^{r/2}},$$

where the constant M does not depend on n, a, b, d.

Proof. Since f is a cusp form of weight r, we have

$$|f(z)| \, (\operatorname{Im} z)^{r/2} < M \tag{11}$$

in the upper half-plane, for some bound $M > 0$. We sketch the proof; see [Apostol 1990] for details. Let $\varphi(z) = |f(z)| \, (\operatorname{Im} z)^{r/2}$ First, $\varphi(z) \to 0$ as $\operatorname{Im} z \to +\infty$, since f decays exponentially with $\operatorname{Im} z$, and therefore faster than any polynomial. By compactness, then, $\varphi(z)$ must be bounded in the fundamental region

$$\left\{ z : \operatorname{Im} z > 0, \ |z| \geq 1, \ \operatorname{Re} z \leq \tfrac{1}{2} \right\}$$

for the action of the modular group Γ on the upper half-plane. But φ is invariant under Γ (basically because $\operatorname{Im} z$ acts like the absolute value of a modular form of weight -2, so the weights cancel out). Thus the value of φ at any point z equals its value at some point in the fundamental domain, and is therefore bounded.

From (11) we can write

$$\left|f\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)\right|\left(\operatorname{Im}\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)\right)^{r/2} < M.$$

Since

$$\left(\operatorname{Im}\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)\right)^{r/2} = |cz+d|^{-r} n^{r/2} (\operatorname{Im} z)^{r/2},$$

we obtain the desired inequality. □

We recall that $r > 0$ for holomorphic cusp forms [Apostol 1990]. Thus, if $\{|h(n)|\} \in \ell^1$, the error term

$$c^{-r} \sum_{n=k+1}^{\infty} h(n) f\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)$$

is clearly absolutely convergent and approaches 0 as $k \to \infty$. Hence, from (7), (8), and (9), we have successively

$$\left(z+\frac{d}{c}\right)^r g_k(z) = c^{-r} \sum_{n=1}^{\infty} h(n) f\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) + O(\epsilon(k)), , \tag{12}$$

for some function $\epsilon(k)$ satisfying $\lim_{k\to\infty} \epsilon(k) = 0$. This leads to

$$\left(z+\frac{d}{c}\right)^r g(z) = c^{-r} \sum_{n=1}^{\infty} h(n) f\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) + O(\epsilon(k)) + O(e^{-2\pi k y}),$$

$$\left(z+\frac{d}{c}\right)^r g(z) = c^{-r} \sum_{n=1}^{\infty} h(n) f\left(n\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) + O(\epsilon(k)).$$

From (12) we obtain, using the Fourier expansion $f(nz) = \sum_{m=1}^{\infty} C(m) e^{2\pi i m n z}$ and absolute convergence to justify rearrangements,

$$\begin{aligned}\left(z+\frac{d}{c}\right)^r g(z) &= c^{-r} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} h(n) C(m) e^{2\pi i m n \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z} + O(\epsilon(k)) \\ &= c^{-r} \sum_{n=1}^{\infty} (h * C)_n e^{2\pi i n \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z} + O(\epsilon(k)).\end{aligned}$$

From (10) we have

$$\left(z + \lim_{c(k)\to\infty} \frac{d}{c}\right)^r g(z) = \lim_{c(k)\to\infty} c^{-r} g\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right),$$

with $c(k) = \operatorname{lcm}(1, 2, 3, \ldots, k)$, which completes the proof of Theorem 1. We note that $g(z)$ is holomorphic in the upper half-plane since $|h(n)|\} \in \ell^1$ and $C(n) = O(n^{r/2})$ result in uniform convergence of the series (10) on compact subsets. □

Corollary 4. *If there exists a cusp form of even weight r over $\mathbb{Q}$ with $C(1) \neq 0$, then the set G of generalized modular forms of weight r and level infinity with rational coefficients has cardinality* $2^{\aleph_0}$:

$$|G| = 2^{\aleph_0}.$$

Proof. This follows from Theorem 1, which implies that, for all $\{h(n)\}$ such that $\{|h(n)|\} \in \ell^1$,

$$g(z) = \sum_{n=1}^{\infty} (h * C)_n e^{2\pi i n z}$$

is a generalized modular form of weight r over $\mathbb{Q}$, assuming that $\{C(n)\}$ is the rational Fourier coefficient sequence of a weight r cusp form with $C(1) \neq 0$.

Now let

$$A = (\mathbb{Q}[0, 1])^{\mathbb{N}} = \bigl\{(a : \mathbb{N} \to \mathbb{Q}[0, 1])\bigr\}$$

be the set of sequences $\{a(n)\}$ with $a(n) \in \mathbb{Q}[0, 1]$ for $n \in \mathbb{N}$. We recall from set theory that $|\mathbb{Q}[0, 1]| = \aleph_0$ and $|(\mathbb{Q}[0, 1])^{\mathbb{N}}| = \aleph_0^{\aleph_0} = 2^{\aleph_0}$ [Jech 1997]. Further, let

$$B = \bigl\{\{h(n)\} \in A : \{|h(n)|\} \in \ell^1\bigr\}$$

be the subset of A consisting of sequences whose sum converges absolutely. We know that $\{a(n)\} \in A$ implies $\{a(n)/n^2\} \in B$, since $|a(n)| \leq 1$ and by the comparison test for series and the absolute convergence of $\sum_{n=1}^{\infty} 1/n^2$. Thus the mapping $\{a(n)\} \to \{a(n)/n^2\}$ defines an injection $\beta : A \to B$.

Next, Theorem 1 implies that there exists an injection $\gamma : B \to G$, which sends a sequence $\{h(n)\} \in B$ to

$$g(z) = \sum_{n=1}^{\infty} (h * C)_n e^{2\pi i n z},$$

with $g(z) \in G$. The composite map $\gamma\beta : A \to G$ thus defines an injection from A to G, as long as $\sum_{n=1}^{\infty} C(n) e^{2\pi i n z}$ is a cusp form of weight r with $C(1) \neq 0$. Hence $|G| \geq 2^{\aleph_0} = |A|$.

At the same time, there is an injection from G into the set S of all formal power series of $q = e^{2\pi i z}$ over $\mathbb{Q}$. This set has the same cardinality as the set $\mathbb{Q}^{\mathbb{N}}$ of maps $\mathbb{N} \to \mathbb{Q}$. Hence $|S| = |\mathbb{Q}^{\mathbb{N}}| = \aleph_0^{\aleph_0} = 2^{\aleph_0}$. We conclude that $|G| \leq 2^{\aleph_0}$. Hence $|G| = 2^{\aleph_0}$. □

We note that Corollary 4 holds for $r = 12$ and all even $r \geq 16$. This is because the standard $\Delta(z)$ function is a cusp form of weight 12 with $C(1) \neq 0$, and for even $r \geq 16$ an example of such a cusp form is $\Delta(z)E_{r-12}(z)$ with $E_{r-12}(z)$ an Eisenstein series.

Proposition 5. *$M_r(\Gamma)$ has cardinality $\aleph_0$ as a vector space over $\mathbb{Q}$.*

Proof. This follows from the result that every entire modular form $f \in M_r(\Gamma)$ is a polynomial of the form [Ono 2004]

$$f = \sum_{4a+6b=r} c_{a,b} G_4^a G_6^b,$$

where G_4 and G_6 are Eisenstein series with integer coefficients, $c_{a,b} \in \mathbb{C}$, and $a, b \in \mathbb{Z}^+$. If f has rational coefficients, then we conclude $c_{a,b} \in \mathbb{Q}$ since G_4 and G_6 have integer coefficients. Algebraically, this implies the following vector space isomorphism over $\mathbb{Q}$:

$$M_r(\Gamma) \cong \mathbb{Q}^{\dim M_r(\Gamma)}.$$

It is a well-known theorem in set theory that $\mathbb{Q}$ is countable, and in general the Cartesian products of any finite number of countable sets is countable [Jech 1997]. Thus, we conclude $M_r(\Gamma)$ over $\mathbb{Q}$ has cardinality $\aleph_0$. These results allow us to gauge the strength of Theorem 1, which generalizes the notion of modularity to encompass a much larger set of holomorphic functions than the classical entire modular forms. □

Acknowledgement

The author thanks the Center for Undergraduate Research and Fellowships and the University Scholars program of the University of Pennsylvania for providing financial support for this work, and Professor J. Haglund for overseeing the project. The author also thanks Professor K. Ono and the referee for their insightful comments, which led to an improved presentation of this paper.

References

[Apostol 1990] T. M. Apostol, *Modular functions and Dirichlet series in number theory*, 2nd ed., Graduate Texts in Mathematics **41**, Springer, New York, 1990. MR 90j:11001 Zbl 0697.10023

[Jech 1997] T. Jech, *Set theory*, 2nd ed., Springer, Berlin, 1997. MR 99b:03061 Zbl 0882.03045

[Ono 2004] K. Ono, *The web of modularity: arithmetic of the coefficients of modular forms and q-series*, CBMS Regional Conference Series in Mathematics **102**, American Mathematical Society, Providence, RI, 2004. MR 2005c:11053 Zbl 1119.11026

[Shimura 2007] G. Shimura, *Elementary Dirichlet series and modular forms*, Springer, New York, 2007. MR 2008g:11001 Zbl 1148.11002

Received: 2010-07-20 Revised: 2011-07-03 Accepted: 2011-08-04

adahaque@sas.upenn.edu *Department of Mathematics, University of Pennsylvania,*
Philadelphia, PA 19104, United States

msp

INVOLVE 5:1(2012)

Induced subgraphs of Johnson graphs

Ramin Naimi and Jeffrey Shaw

(Communicated by Jerrold Griggs)

The Johnson graph $J(n, N)$ is defined as the graph whose vertices are the n-subsets of the set $\{1, 2, \dots, N\}$, where two vertices are adjacent if they share exactly $n-1$ elements. Unlike Johnson graphs, induced subgraphs of Johnson graphs (JIS for short) do not seem to have been studied before. We give some necessary conditions and some sufficient conditions for a graph to be JIS, including: in a JIS graph, any two maximal cliques share at most two vertices; all trees, cycles, and complete graphs are JIS; disjoint unions and Cartesian products of JIS graphs are JIS; every JIS graph of order n is an induced subgraph of $J(m, 2n)$ for some $m \leq n$. This last result gives an algorithm for deciding if a graph is JIS. We also show that all JIS graphs are edge move distance graphs, but not vice versa.

1. Introduction

We work with finite, simple graphs. Let $F = \{S_1, \dots, S_m\}$ be a family of finite sets. The *intersection graph* of F, denoted $\Omega(F)$, is the graph whose vertices are the elements of F, where two vertices S_i and S_j, $i \neq j$, are adjacent if they share at least one element. More generally, for a fixed positive integer p, the *p-intersection graph* of F, denoted $\Omega_p(F)$, is the graph whose vertices are the elements of F, where two vertices are adjacent if they share at least p elements. (Thus $\Omega_p(F)$ is a subgraph of $\Omega_1(F) = \Omega(F)$.) McKee and McMorris [1999] give an extensive and excellent survey of intersection graphs, which also includes a section on p-intersection graphs. Here we narrow attention to p-intersection graphs of families of $(p+1)$-sets, so that two vertices S_i and S_j are adjacent if $|S_i \cap S_j| = |S_i| - 1 = |S_j| - 1$, i.e., S_i and S_j differ by exactly one element.

Another way to view these graphs is as induced subgraphs of Johnson graphs. Given positive natural numbers $n \leq N$, the *Johnson graph* $J(n, N)$ is defined as the graph whose vertices are the n-subsets of the set $\{1, 2, \dots, N\}$, where two vertices are adjacent if they share exactly $n-1$ elements. Hence a graph G is isomorphic

MSC2000: 05C62.

Keywords: Johnson graph, intersection graph, distance graph.

to an induced subgraph of a Johnson graph if and only if it is possible to assign, for some fixed n, an n-set S_v to each vertex v of G such that distinct vertices have distinct corresponding sets, and vertices v and w are adjacent if and only if S_v and S_w share exactly $n-1$ elements. When this happens, we say the family of n-sets $F = \{S_v : v \in V(G)\}$ *realizes* G as an induced subgraph of a Johnson graph, which we abbreviate by saying G is *JIS*. Thus, F realizes G as a JIS graph if and only if G is isomorphic to $\Omega_{n-1}(F)$, which in turn is isomorphic to an induced subgraph of $J(n, N)$, where $N = \left|\bigcup_{S \in F} S\right|$.

Although there is a considerable amount of literature written on Johnson graphs, we have not been able to find any on their induced subgraphs. It would be desirable to obtain "nice" necessary and sufficient conditions for when a graph is JIS. In this paper, we only give some necessary conditions and some sufficient conditions.

A *clique* in a graph G is a complete subgraph of G. A clique L in G is called a *maximal clique*, or a *maxclique* for short, if there is no larger clique $L' \subseteq G$ that contains L. In Section 2 we describe how the maxcliques of a graph play a role in whether or not it is JIS. In particular, Proposition 2(1) states that any two distinct maxcliques in a JIS graph can share at most two vertices. It follows, for example, that the graph "K_5 minus one edge" is not JIS, since it contains two maximal 4-cliques that share three vertices.

The conditions given in Section 2 are necessary, but not sufficient, for a graph to be JIS. In Section 3 we show that the complete bipartite graph $K_{2,3}$, as well as a few other graphs, satisfy all these necessary conditions but are not JIS. In Section 3 we also give some sufficient conditions for a graph to be JIS, including the following:

- All complete graphs and all cycles are JIS.
- A graph is JIS if and only if all its connected components are JIS.
- The Cartesian product of two JIS graphs is JIS.

Despite not having a "nice" characterization of JIS graphs, for any graph G the question "Is G JIS?" is decidable; this follows from Theorem 10, which says that every JIS graph of order n is isomorphic, for some $m \leq n$, to an induced subgraph of the Johnson graph $J(m, 2n)$. In other words, every JIS graph of order n can, for some $m \leq n$, be realized by m-subsets of $\{1, 2, \ldots, 2n\}$. This gives us a simple (albeit slow) algorithm for determining if a graph G is JIS: Do an exhaustive search among all n-families of m-subsets of $\{1, \ldots, 2n\}$, where n is the order of G and $m \leq n$, to see if any of them realizes G as a JIS graph.

The *p-intersection number* of a graph G is defined as the smallest k such that G is isomorphic to the p-intersection graph of a family of subsets of $\{1, \ldots, k\}$ ([McKee and McMorris 1999], p. 91). Thus, an immediate corollary of Theorem 10 is that every JIS graph of order n has, for some $m \leq n$, $(m-1)$-intersection number at most $2n$.

In the final section of this paper we discuss edge move distance graphs and their relationship to JIS graphs.

2. Maxcliques in JIS Graphs

Given n-sets $S_1, \ldots, S_k$ with $n \geq 1$ and $k \geq 2$, we say they share an *immediate subset* if $\left|\bigcap_{i=1}^{k} S_i\right| = n-1$. Similarly, $S_1, \ldots, S_k$ share an *immediate superset* if $\left|\bigcup_{i=1}^{k} S_i\right| = n+1$. Observe that for $k = 2$, S_1 and S_2 share an immediate subset if and only if they share an immediate superset: $|S_1 \cup S_2| = |S_1| + |S_2| - |S_1 \cap S_2| = 2n - |S_1 \cap S_2|$; hence $|S_1 \cup S_2| = n+1$ if and only if $|S_1 \cap S_2| = n-1$. We begin with the following elementary result on realizations of complete graphs as JIS graphs.

Lemma 1. *Let $S_1, \ldots, S_k$ be n-sets that pairwise share an immediate subset, where $n \geq 1$ and $k \geq 3$. Then $S_1, \ldots, S_k$ share an immediate subset or an immediate superset, but not both.*

Proof. We first show that for $k \geq 3$, if $S_1, \ldots, S_k$ share an immediate subset, then they do not share an immediate superset. Suppose $T = S_1 \cap \cdots \cap S_k$ has $n-1$ elements. Then, for each i, $S_i \setminus T$ has exactly one element, x_i. For all $j \neq i$, $x_i \notin S_j$ since $S_i \neq S_j$. It follows that $S_1 \cup \cdots \cup S_k$ has at least $n - 1 + k \geq n + 2$ elements, since $k \geq 3$. Thus $S_1, \ldots, S_k$ do not share an immediate superset.

Now suppose $S_1, \ldots, S_k$ pairwise share an immediate subset. We use induction on k to prove that they share an immediate subset or an immediate superset.

Assume $k = 3$. Let $T = S_1 \cap S_2$. If $T \subset S_3$, then $|S_1 \cap S_2 \cap S_3| = |T| = n-1$, and we're done. So assume $T \not\subset S_3$. Note that $|S_1 \setminus T| = |S_2 \setminus T| = 1$. Hence, for S_3 to share $n-1$ elements with each of S_1 and S_2, it must contain an $(n-2)$-subset of T, as well as $S_1 \setminus T$ and $S_2 \setminus T$, and no other elements. It follows that $|S_1 \cup S_2 \cup S_3| = n+1$, as desired.

Now assume $k \geq 4$. Then, by our induction hypothesis, $S_1, \ldots, S_{k-1}$ share an immediate subset or an immediate superset; and similarly for $S_2, \ldots, S_k$. We have four cases:

Case 1: $S_1, \ldots, S_{k-1}$ share an immediate subset and $S_2, \ldots, S_k$ share an immediate subset. Then $S_1, \ldots, S_k$ share $S_2 \cap S_3$ as an immediate subset.

Case 2: $S_1, \ldots, S_{k-1}$ share an immediate superset and $S_2, \ldots, S_k$ share an immediate superset. Then $S_1, \ldots, S_k$ share $S_2 \cup S_3$ as an immediate superset.

Case 3: $S_1, \ldots, S_{k-1}$ share an immediate subset and $S_2, \ldots, S_k$ share an immediate superset. Let $T = S_1 \cap \cdots \cap S_{k-1}$. Then, for $1 \leq i \leq k-1$, $S_i \setminus T$ has exactly one element, x_i; and, for $1 \leq j \leq k-1$ with $j \neq i$, $x_i \notin S_j$ since $S_i \neq S_j$. Since $|S_2 \cup \cdots \cup S_k| = n+1 = |S_2 \cup S_3|$, S_k is a proper subset of $S_2 \cup S_3 = T \cup \{x_2, x_3\}$. And since S_2, S_3, S_k share an immediate superset, they do not share an immediate subset; hence $T \not\subset S_k$. This implies that $x_2, x_3 \in S_k$ since S_k has n elements and

$T \cup \{x_2, x_3\}$ has $n + 1$ elements. But $x_2, x_3 \notin S_1$, so $|S_1 \cap S_k| < n - 1$, which contradicts the hypothesis of the lemma.

Case 4: $S_1, \dots, S_{k-1}$ share an immediate superset and $S_2, \dots, S_k$ share an immediate subset. This case is similar to Case 3. □

We now use Lemma 1 to establish restrictions on how maxcliques in a JIS graph can intersect or connect to each other by edges.

Proposition 2. *Suppose G is JIS and L and L' are distinct maxcliques in G.*

(1) *L and L' share at most two vertices.*

(2) *If L and L' share exactly two vertices, then no vertex in $V(L) \setminus V(L')$ is adjacent to a vertex in $V(L') \setminus V(L)$.*

(3) *If L and L' share exactly one vertex, then each vertex in either of the two sets $V(L) \setminus V(L')$ and $V(L') \setminus V(L)$ is adjacent to at most one vertex in the other set.*

Proof. Let $\{S_v : v \in V(G)\}$ be a family of n-sets that realizes G as a JIS graph.

(1) Suppose towards contradiction that L and L' are distinct maxcliques that share three (or more) vertices, u, v, and w. Let x be a vertex of L not in L', and x' a vertex of L' not in L; x and x' exist since L and L' are distinct and maximal. Then, by Lemma 1, the sets S_x, S_u, S_v, and S_w share an immediate subset or an immediate superset. Similarly for $S_{x'}$, S_u, S_v, and S_w. But S_u, S_v, and S_w cannot share both an immediate subset and an immediate superset. It follows that S_x and $S_{x'}$ share an immediate subset or an immediate superset, which implies that x and x' are adjacent. Hence every vertex of L is adjacent to every vertex of L', but this contradicts the assumption that L is a maxclique in G.

(2) Let L and L' be distinct maxcliques that share exactly two vertices, v and w. Suppose towards contradiction that there exist adjacent vertices $x \in V(L) \setminus V(L')$ and $x' \in V(L') \setminus V(L)$. Then the induced subgraph of G containing $\{x, x', v, w\}$ is a 4-clique. Let L'' be the maxclique that contains this 4-clique. Then L'' is distinct from L and shares at least three vertices with it. This contradicts (1).

(3) The proof is similar to the proof of (2). Let L and L' be distinct maxcliques that share exactly one vertex, v. Suppose towards contradiction that there exist vertices $x \in V(L) \setminus V(L')$ and $x', y' \in V(L') \setminus V(L)$ with x adjacent to x' and y'. Then the induced subgraph of G containing $\{x, x', y', v\}$ is a 4-clique, and the maxclique that contains this 4-clique is distinct from L' and shares at least three vertices with it. This contradicts (1). □

Proposition 3. *Suppose $L_1, \dots, L_k$, where k is odd and at least* 3, *are distinct maxcliques in a graph G such that L_i shares exactly two vertices with L_{i+1} for $1 \le i \le k - 1$, and L_k shares exactly two vertices with L_1; then G is not JIS.*

Proof. In the following, L_{i+1} refers to L_1 whenever $i=k$. Suppose towards contradiction that G is realized as a JIS graph by a family of n-sets. Note that each L_i has at least three vertices, since otherwise it would not be distinct from L_{i+1}. Hence, by Lemma 1, we can label each L_i as either "sub" or "super" according to whether the n-sets assigned to its vertices share an immediate subset or an immediate superset. Then, since k is odd, there exists a j such that L_j and L_{j+1} have the same label. Now, L_j and L_{j+1} share two vertices; therefore the n-sets assigned to their vertices must all share the same immediate subset or immediate superset, which makes all vertices in L_j adjacent to those in L_{j+1}, giving a contradiction. □

An equivalent way of stating the above result is: One can label every maxclique in a JIS graph with a $+$ or $-$ (or any two symbols) in such a way that any two maxcliques that share two vertices have distinct labels.

3. Miscellaneous JIS and non-JIS graphs

In this section we give some sufficient conditions for when a graph is JIS. We also describe some graphs that satisfy all the conditions listed in the results of the previous section as necessary for a graph to be JIS, but are not JIS.

Proposition 4. *All complete graphs and all cycles are JIS.*

Proof. For each n, K_n is realized as a JIS graph by the 1-sets $\{1\}, \{2\}, \dots, \{n\}$. For each $n \geq 3$, the n-cycle is realized as a JIS graph by the 2-sets $\{1,2\}$, $\{2,3\}$, $\dots$, $\{n-1,n\}$, $\{n,1\}$. □

We define the *n-core* of a graph G as the graph obtained by recursively removing all vertices of degree less than n until there are none left.

Proposition 5. *A graph is JIS if and only if its* 2*-core is JIS.*

Proof. Suppose G is obtained from a graph G' by removing exactly one vertex, w, which has degree 0 or 1. By induction, it is enough to show that G is JIS if and only if G' is JIS. Clearly, if G' is JIS, then so is G, since any induced subgraph of a JIS graph is JIS. To prove the converse, suppose G is JIS. Let $\{S_x : x \in V(G)\}$ be n-sets that realize G as a JIS graph. Pick distinct a and b that are not in any of the sets S_x. For each $x \in V(G)$, let $S'_x = S_x \cup \{a\}$. Let $S'_w = S_v \cup \{b\}$, where $v \in V(G')$ is arbitrary if w has degree 0, and v is adjacent to w if w has degree 1. Then $\{S'_x : x \in V(G')\}$ are $(n+1)$-sets that realize G' as a JIS graph, as desired. □

It follows as a trivial corollary that all trees are JIS.

Proposition 6. *A graph is JIS if and only if all its connected components are JIS.*

Proof. One direction is trivial: every induced subgraph of a JIS graph, and in particular every connected component of it, is JIS. We prove the converse by induction on the number of components of G.

Base step: Suppose that G has two components, G_i, $i = 1, 2$, each realized as a JIS graph by a family of sets F_i. We can assume without loss of generality that each set in F_1 is disjoint from each set in F_2.

We would like each set in F_1 to have the same size as each set in F_2, in order to obtain $F_1 \cup F_2$ as a family that realizes G as a JIS graph. If this is not already so, we proceed as follows. Let m_i denote the number of elements in each set in F_i. We can assume $n_1 > n_2$. Now add the first $n_1 - n_2$ elements of the first set in F_1 to every set in F_2.

Once the sets in the two families all have the same size, we must make sure that sets corresponding to vertices in different components of G do not share immediate subsets. This will automatically be true for sets that had two or more elements before any extra elements were added to them (since we started with the sets in F_1 disjoint from those in F_2), but not for singletons. We remedy this by adding, for each i, an element e_i to every set in F_i, where e_1 and e_2 are distinct elements not already in any set in any F_i. It is now easy to verify that $F_1 \cup F_2$ realizes G as a JIS graph.

The inductive step follows trivially from the base step. □

Proposition 7. *The Cartesian product of two JIS graphs is JIS.*

Proof. Let G and G' be JIS graphs that are realized, respectively, by sets $\{S_x : x \in V(G)\}$ and $\{S'_{x'} : x' \in V(G')\}$. We can assume without loss of generality that every S_x is disjoint from every $S'_{x'}$.

For each vertex $v = (x, x') \in V(G \times G')$, let $T_v = S_x \cup S'_{x'}$. By definition, two vertices $v = (x, x')$ and $w = (y, y')$ of $G \times G'$ are adjacent if and only if $x = x'$ and y is adjacent to y' or $y = y'$ and x is adjacent to x'. Thus, T_v and T_w share an immediate subset if and only if v and w are adjacent. Hence the sets $\{T_v : v \in G \times G'\}$ realize $G \times G'$ as a JIS graph. □

Proposition 8. *The complete bipartite graph $K_{2,3}$ is not JIS.*

Proof. Label the two degree-3 vertices of $K_{2,3}$ as v and w, and the three degree-2 vertices as x, y, and z, as in Figure 1. Suppose towards contradiction that there exists a family of n-sets $\{S_u : u \in V(K_{2,3})\}$ that realizes $K_{2,3}$ as a JIS graph. Since v and w have distance two (where *distance* is the number of edges in the shortest

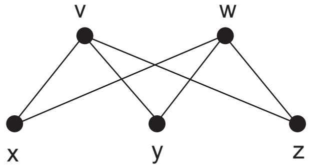


Figure 1. $K_{2,3}$ with labeled vertices.

path joining the two vertices), S_v and S_w must share exactly $n-2$ elements (this does not work for distance ≥ 3; it works only for distance ≤ 2). Let $T = S_v \cap S_w$. Then, since each of x, y, and z is adjacent to both v and w, S_x, S_y, and S_z must each contain T as a subset. Therefore, by subtracting T from every S_u, $u \in V(K_{2,3})$, we get a family of 2-sets that realizes $K_{2,3}$. Hence we will assume that every S_u has exactly two elements. It follows that S_v and S_w are disjoint; and S_x, S_y, and S_z are pairwise disjoint and each shares exactly one element with each of S_v and S_w.

So, without loss of generality, $S_v = \{1, 2\}$, and $S_w = \{3, 4\}$. Therefore, again without loss of generality, $S_x = \{1, 3\}$, and $S_y = \{2, 4\}$. And there is nothing left for S_z. □

The graph $K_{2,3}$ can be thought of as two 4-cycles that share three vertices. So one may wonder whether the graph θ_n consisting of two n-cycles that share $n-1$ vertices is also not JIS. It turns out that θ_n is not JIS only for $n = 4$ and $n = 5$. The proof that θ_5 is not JIS is very similar to the proof that $K_{2,3}$ is not JIS, and we therefore omit it. The proof that θ_n is JIS for $n \geq 6$ is a straightforward construction, which we also omit.

One may also wonder whether $K_{2,3}$ becomes JIS if an edge is added to it. There are, up to isomorphism, two ways to add an edge to $K_{2,3}$: add an edge that connects the two degree-3 vertices; or add an edge that connects two of the three degree-2 vertices. It turns out that neither of these two graphs is JIS. The proof that the former graph is not JIS follows immediately from Proposition 3. The proof that the latter graph (which we call Δ_2) is not JIS is given below in Proposition 9.

The graphs Δ_i depicted in Figure 2 have the following pattern (ignore the vertex labels and the $+$ and $-$ signs for now; they are used later): Δ_i consists of a chain of i "consecutively adjacent" triangles, plus one vertex which is connected to the two vertices of degree 2 in the triangle chain. It turns out that, like $K_{2,3}$, Δ_2, Δ_4, and Δ_6 satisfy the necessary conditions in the results of the previous sections for being JIS, but are not JIS; Δ_3 and Δ_5, however, are JIS. We prove these claims below, except for Δ_6: its proof is similar to that of Δ_2 and Δ_4, but is more tedious, and in our opinion not worth being included here. We did not check which Δ_i are JIS for $i \geq 7$, but, from the pattern for $i \leq 6$, it seems that:

Conjecture. Δ_i is JIS if and only if i is odd.

Proposition 9. (*i*) *The graphs* Δ_2 *and* Δ_4 *are not JIS.* (*ii*) *The graphs* Δ_3 *and* Δ_5 *are JIS.*

Remark. As mentioned above, Δ_2 is isomorphic to $K_{2,3}$ plus an edge that connects two of its three degree-2 vertices. Because of this, the proof that $K_{2,3}$ is not JIS can be easily modified to prove that Δ_2 is not JIS. However, we give a different proof below, one that can be naturally extended to also prove that Δ_4 (and Δ_6) is not JIS.

Proof. Label the vertices of Δ_2 as $v_1, \dots, v_5$, as in Figure 2. The $+$ and $-$ signs will be explained shortly. Suppose, towards contradiction, that Δ_2 can be realized as a JIS graph by sets $S_1, \dots, S_5$ (for simplicity, we write S_i instead of S_{v_i}). Each of the two triangles in Δ_2 is a maxclique. Thus, by Lemma 1, S_1, S_2, and S_3 must share an immediate subset or an immediate superset; similarly for S_2, S_3, and S_4. Furthermore, S_1, S_2, and S_3 share an immediate subset if and only if S_2, S_3, and S_4 share an immediate superset, because: if S_1, S_2, and S_3 share an immediate subset and S_2, S_3, and S_4 also share an immediate subset, then S_1 and S_4 must share $S_2 \cap S_3$ as an immediate subset, but this contradicts the fact that v_1 and v_4 are not adjacent; and if S_1, S_2, and S_3 share an immediate superset and S_2, S_3, and S_4 also share an immediate superset, then S_1 and S_4 must share $S_2 \cup S_3$ as an immediate superset, which implies that they also share an immediate subset, again contradicting the fact that v_1 and v_4 are not adjacent.

Thus, without loss of generality, we will assume that S_1, S_2, and S_3 share an immediate subset. This is indicated in Figure 2 by the $-$ sign; the $+$ signs indicate immediate supersets. So we will assume that $S_1 = \{1, 2, 3, 4\}$, $S_2 = \{1, 2, 3, 5\}$, and $S_3 = \{1, 2, 3, 6\}$; we explain in the next paragraph why there is no loss of generality in assuming that S_i are 4-sets (as opposed to larger sets). To make the notation more compact, we will drop the commas and the braces from each set; e.g., $S_1 = 1234$. Then S_4 must be a 4-subset of $S_2 \cup S_3 = 12356$. Since S_1 and S_4 have no immediate subset, we can without loss of generality assume that $S_4 = 2356$. Now, S_5 must differ by exactly one element from each of S_1 and S_4. The only possibilities are 1235, 1236, 2345, and 2346. But the first two are equal to S_2 and S_3 respectively; and the last two differ from S_2 and S_3 respectively by exactly one element, which is not allowed since v_5 is adjacent to neither v_2 nor v_3. Thus we have a contradiction, as desired.

Note that by assuming that all S_i are 4-sets, we ended up with all of them sharing the two elements 2 and 3. If we instead assumed that S_i were n-sets with $n \geq 5$,

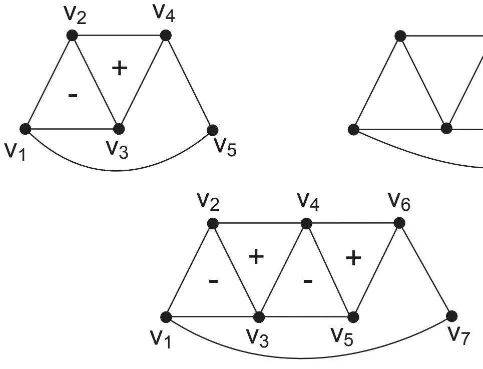


Figure 2. Δ_2, Δ_3, and Δ_4, with vertices labeled in Δ_2 and Δ_4.

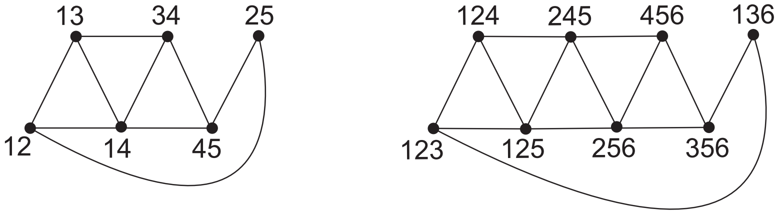


Figure 3. Δ_3 (left) and Δ_5 (right) realized as JIS graphs.

the proof would remain the same except that we would end up with all S_i sharing more than two elements. Hence there is no loss of generality in assuming that S_i are 4-sets (in fact, this shows that we could even assume they are 2-sets).

To prove that Δ_4 is not JIS, we start with the same assumptions that S_1, S_2, and S_3 share an immediate subset, S_2, S_3, and S_4 share an immediate superset, and $S_1 = 1234$, $S_2 = 1235$, $S_3 = 1236$, and $S_4 = 2356$. Now, S_3, S_4, and S_5 must share an immediate subset. So S_5 must contain $S_3 \cap S_4 = 236$. Since v_5 is adjacent to neither v_1 nor v_2, S_5 can contain neither 1 nor 4 nor 5. Hence, without loss of generality, $S_5 = 2367$. Continuing, S_4, S_5, and S_6 must share an immediate superset. So S_6 must be a 4-subset of $S_4 \cup S_5 = 23567$; i.e., we must drop one element from 23567 to get S_6. Dropping 5 or 7 gets us back to S_4 and S_5; hence we must drop 2, 3, or 6. The roles of 2 and 3 have been identical so far; so, without loss of generality, we must drop 2 or 6; so $S_6 = 2357$ or 3567. The former is not possible since v_6 and v_2 are not adjacent. And the latter is ruled out by noticing that 3567 differs from $S_1 = 1234$ by three elements, which contradicts the fact that v_6 and v_1 have distance two[1]. Thus we have reached a contradiction, as desired.

Part (ii) of the proposition is proved in Figure 3, which shows sets that realize Δ_3 and Δ_5 as JIS graphs. For the sake of compactness, braces and commas are omitted from the sets. □

We end this section with the following definition and question. Let G be a JIS graph, and suppose $F = \{S_u : u \in V(G)\}$ realizes G as a JIS graph. We define the *F-distance* between two vertices v and w of G to be $d_F(v, w) = |S_v \setminus S_w|$. It is easy to show this distance function is indeed a metric. The *JIS-diameter* of G is defined as

$$\max_{v,w \in V(G)} \min_{F} \{d_F(v, w)\}$$

where the minimum is taken over all families F that realize G as a JIS graph.

Question. Do there exist JIS graphs with arbitrarily large JIS-diameter?

[1] Note that $\Delta_4 - v_7$ *is* JIS, with S_1 and S_6 differing in three elements. We will refer back to this point at the very end of this section.

From the proof of Proposition 9 and the footnote in it, it follows that Δ_4 minus the degree-2 vertex v_7 has JIS-diameter 3: $S_1 = 1234$, $S_2 = 1235$, $S_3 = 1236$, $S_4 = 2356$, $S_5 = 2367$, and $S_6 = 3567$, i.e., v_1 and v_6 have F-distance 3.

4. An algorithm for recognizing JIS graphs

As mentioned in the introduction, the following theorem provides for an algorithm for deciding if a graph is JIS by doing a bounded exhaustive search.

Theorem 10. *Every JIS graph of order n is isomorphic, for some $m \leq n$, to an induced subgraph of the Johnson graph $J(m, 2n)$.*

Proof. Let G be a JIS graph of order n with c connected components.

Case 1. Assume $c = 1$, i.e., G is connected. In this case we will prove a slightly stronger result, which we will use in the proof of Case 2:

G is isomorphic, for some $m \leq n$, to an induced subgraph of $J(m, 2n-1)$.

The case $n = 1$ is trivial; so we assume $n \geq 2$. Since G is connected, there exists an ordering $v_1, v_2, \ldots, v_n$ of the vertices of G such that for each $i \geq 2$, v_i is adjacent to at least one of $v_1, \ldots, v_{i-1}$. Since G is JIS, for some $k \geq 1$ there exist k-sets $\{S_1, \ldots, S_n\}$ that realize G as a JIS graph, where S_i corresponds to the vertex v_i. Since v_1 and v_2 are adjacent, $|S_1 \cap S_2| = k - 1$. Since v_3 is adjacent to at least one of v_1 and v_2, $|S_1 \cap S_2 \cap S_3| \geq k - 2$. Continuing this way, we see that $|S_1 \cap \cdots \cap S_n| \geq k - (n - 1)$. Let

$$S'_i = S_i \setminus (S_1 \cap \cdots \cap S_n)$$

for $1 \leq i \leq n$. Then for all i, $|S'_i| = m$ where $m \leq k - (k - (n - 1)) = n - 1$, and it is easily verified that the family of sets $\{S'_1, \ldots, S'_n\}$ realizes G as a JIS graph.

Now, since v_1 and v_2 are adjacent, $|S'_1 \cup S'_2| = m + 1$. Since v_3 is adjacent to at least one of v_1 and v_2, $|S'_1 \cup S'_2 \cup S'_3| \leq m + 2$. Continuing this way, we see that $|S'_1 \cup \cdots \cup S'_n| \leq m + n - 1 \leq 2n - 2$, which implies G is an induced subgraph of $J(m, 2n-1)$, $m \leq n-1$. (Note: we proved the inequalities $|S'_1 \cup \cdots \cup S'_n| \leq 2n-2$ and $m \leq n - 1$ only for $n \geq 2$, not for $n = 1$.)

Case 2. Assume $c \geq 2$. Let n_i be the order of the ith component of G. Then, by Case 1 above, for each i there is a family F_i of m_i-sets, $m_i \leq n_i$, that realizes the ith component of G as a JIS graph, such that the union of the sets in F_i has at most $2n_i - 1$ elements. Thus $\bigcup F_i$ has at most $2n - c$ elements.

We can assume $m_1 \geq m_i$ for all i. We can also assume that for all $i \neq j$, every set in the family F_i is disjoint from every set in F_j. To make all sets in all the families have the same size, for each i such that $m_1 > m_i$ we add the first $m_1 - m_i$ elements of the first set in F_1 to every set in F_i. After adding these extra elements, we must make sure that sets corresponding to vertices in different components of G do not share immediate subsets. This will automatically be true for sets that had two or more elements before the extra elements were added, but not for singletons. We remedy this by adding, for each i, an element e_i to every set in F_i, where $e_1, \dots, e_c$ are distinct elements not already in any set in any F_i. Let $F = \bigcup F_i$. Then G is realized as a JIS graph by F, which is a family of (m_1+1)-sets whose union has at most $2n - c + c = 2n$ elements, where $m_1 + 1 \leq n_1 + 1 \leq n$. Thus G is an induced subgraph of $J(m, 2n)$ where $m = m_1 + 1 \leq n$. □

Remark. It is not difficult to modify the above proof in Case 1 to show that if G is connected, then it is an induced subgraph of $J(n, 2n)$. It would be interesting to see for which graphs the bounds n and $2n$ can be lowered. Note that if G consists of exactly $n \geq 2$ vertices of degree zero, then the bound $2n$ is optimal.

5. Edge move distance graphs and JIS graphs

Since the 1970s many authors have written on various metrics defined on sets of graphs; see, for instance, [Benadé et al. 1991; Chartrand et al. 1997; 1990; Deza and Deza 2009; Johnson 1987; Kaden 1983; Zelinka 1985]. Among them are edge move, edge rotation, edge jump, and edge slide distances. In general, given a metric d on a set of graphs $S = \{G_1, \dots, G_k\}$, the *distance graph* of S, denoted $D_d(S)$, has S as its vertex set, where two vertices G_i and G_j are adjacent if $d(G_i, G_j) = 1$. We will see shortly that distance graphs associated with the edge move metric are closely related to JIS graphs.

An *edge move* on a graph G consists of removing one edge from and adding a new edge to G, without changing its vertex set $V(G)$; i.e., one edge is “moved to a new position.” The *edge move distance* $d_m(G, H)$ between two graphs G and H is defined as the fewest number of edge moves necessary to transform G into H, up to isomorphism. Note that for $d_m(G, H)$ to be defined, G and H must have the same order and the same size. It is easy to verify that d_m is a metric on any set of graphs of given order and size. Given a set S of graphs of the same order and size, the *edge move distance graph* of S, $D_m(S)$, is the graph whose vertices are the elements of S, where two vertices are adjacent if their edge move distance is one. When we say a graph is an edge move distance graph we mean it is isomorphic to one.

The connection between JIS graphs and edge move distance graphs can be seen by focusing on edge sets. Let G and H be graphs of the same order and size, with n edges each. If the edge sets $E(G)$ and $E(H)$ share exactly $n-1$ elements, then G and H have edge move distance one. Conversely, if G and H have edge move distance one, then their vertices can be labeled such that $E(G)$ and $E(H)$ share exactly $n-1$ elements. At first glance, this might seem to suggest that a graph is JIS if and only if it is isomorphic to an edge move distance graphs. We will show, however, that only half (one direction) of this statement is true.

Proposition 11. *Every JIS graph is an edge move distance graph.*

Proof. Let G be realized as a JIS graph by a family of n-sets $\{S_v : v \in V(G)\}$. We will construct a graph G_v for each $v \in V(G)$ such that $d_m(G_v, G_w) = 1$ if and only if S_v and S_w share an immediate subset.

We can assume that each S_v consists of positive integers. Let

$$k = 1 + \max\{i \in S_v : v \in V(G)\},$$

and let P be a path of length $2k$. Denote the vertices of P by $p_0, p_1, \dots, p_{2k}$. For each $v \in V(G)$, we let G_v be the graph consisting of P plus the edges $p_i p_{2k-i}$ for all $i \in S_v$. Then it is easily verified that for $v \neq w$, G_v is not isomorphic to G_w, and $d_m(G_v, G_w) = 1$ if and only if S_v and S_w share an immediate subset. Therefore G is isomorphic to the edge move distance graph $D_m(\{G_v : v \in V(G)\})$. □

The converse is not true. The reason is that the number of edges shared by the edge sets of two graphs depends on how their vertices are labeled, whereas edge move distance is measured up to graph isomorphism.

Proposition 12. *The graph obtained by removing one edge from the complete graph K_n, where $n \geq 5$, is an edge move distance graph but is not JIS.*

Proof. Fix $n \geq 5$, and let H be the graph obtained by removing one edge from K_n. Then H contains two maximal $(n-1)$-cliques which share $n-2$ vertices. Hence, by Proposition 2(1), H is not JIS.

To show that H is an edge move distance graph, we construct a set of graphs $S = \{Q_1, Q_2, \dots, Q_n\}$ such that $H \simeq D_m(S)$. For $1 \leq i \leq n$, Q_i has $n+2$ vertices: $V(Q_i) = \{v_1, v_2, \dots, v_{n+2}\}$. For $1 \leq i \leq n-1$, we have

$$E(Q_i) = \{v_k v_{k+1} : 1 \leq k \leq n\} \cup \{v_{n-1}v_{n+1}, v_i v_{n+2}\};$$

and $E(Q_n) = (E(Q_1) \cup \{v_1 v_{n-2}\}) \setminus \{v_{n-2}v_{n-1}\}$.

Then one readily verifies for all $i \neq j$ except when $\{i, j\} = \{n-1, n\}$ that Q_i and Q_j have edge move distance one. Thus H is an edge move distance graph. □

Figures 4 and 5 show some of the Q_i in the case $n = 6$.

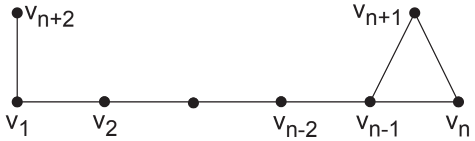


Figure 4. Q_1 for $n = 6$.

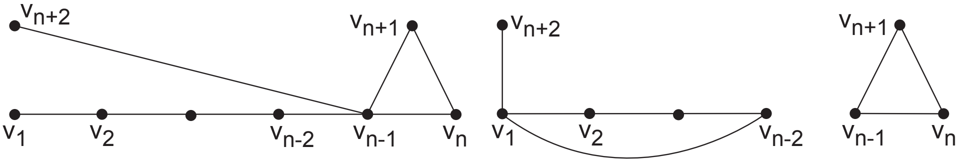


Figure 5. Q_{n-1} (left) and Q_n (right) for $n = 6$.

Acknowledgments

We thank Terry A. McKee of Wright State University for bringing to our attention that the graphs we were studying are related to Johnson graphs. Naimi thanks Caltech for its hospitality while he did part of this work there during his sabbatical leave. Shaw thanks the Undergraduate Research Center of Occidental College for providing support to do this work as part of an undergraduate summer research project. We also thank the (anonymous) referee for helpful suggestions.

References

[Benadé et al. 1991] G. Benadé, W. Goddard, T. A. McKee, and P. A. Winter, "On distances between isomorphism classes of graphs", *Math. Bohem.* **116**:2 (1991), 160–169. MR 92g:05173 Zbl 0753.05067

[Chartrand et al. 1990] G. Chartrand, W. Goddard, M. A. Henning, L. Lesniak, H. C. Swart, and C. E. Wall, "Which graphs are distance graphs?", *Ars Combin.* **29**:A (1990), 225–232. MR 97f:05156 Zbl 0716.05028

[Chartrand et al. 1997] G. Chartrand, H. Gavlas, H. Hevia, and M. A. Johnson, "Rotation and jump distances between graphs", *Discuss. Math. Graph Theory* **17**:2 (1997), 285–300. MR 99f:05084 Zbl 0902.05022

[Deza and Deza 2009] M. M. Deza and E. Deza, *Encyclopedia of distances*, Springer, Berlin, 2009. MR 2011b:51001 Zbl 1167.51001

[Johnson 1987] M. A. Johnson, "An ordering of some metrics defined on the space of graphs", *Czechoslovak Math. J.* **37**:1 (1987), 75–85. MR 88d:05158 Zbl 0641.05027

[Kaden 1983] F. Kaden, "Graph metrics and distance-graphs", pp. 145–158 in *Graphs and other combinatorial topics* (Prague, 1982), edited by M. Fiedler, Teubner-Texte in Mathematik **59**, Teubner, Leipzig, 1983. MR 85e:05060 Zbl 0528.05055

[McKee and McMorris 1999] T. A. McKee and F. R. McMorris, *Topics in intersection graph theory*, SIAM, Philadelphia, 1999. MR 2000e:05001 Zbl 0945.05003

[Zelinka 1985] B. Zelinka, "Comparison of various distances between isomorphism classes of graphs", *Časopis Pěst. Mat.* **110**:3 (1985), 289–293. MR 87c:05114 Zbl 0579.05056

Received: 2010-08-04 Revised: 2011-07-01 Accepted: 2011-07-09

rnaimi@oxy.edu
Department of Mathematics, Occidental College, 1600
Campus Road, Los Angeles, CA 90041-3314, United States

jeffreyeshaw@gmail.com
Department of Mathematics, Occidental College, 1600
Campus Road, Los Angeles, CA 90041-3314, United States

mathematical sciences publishers msp

Multiscale adaptively weighted least squares finite element methods for convection-dominated PDEs

Bridget Kraynik, Yifei Sun and Chad R. Westphal

(Communicated by John Baxley)

We consider a weighted least squares finite element approach to solving convection-dominated elliptic partial differential equations, which are difficult to approximate numerically due to the formation of boundary layers. The new approach uses adaptive mesh refinement in conjunction with an iterative process that adaptively adjusts the least squares functional norm. Numerical results show improved convergence of our strategy over a standard nonweighted approach. We also apply our strategy to the steady Navier–Stokes equations.

1. Introduction

In this paper we consider numerically approximating solutions to the convection-diffusion partial differential equation

$$\begin{cases} -\varepsilon\Delta u + \boldsymbol{b}\cdot\nabla u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{1}$$

Here, $u = u(x, y)$ is the solution, ∇u and Δu are the gradient and Laplacian of u, $\partial\Omega$ is the boundary of domain Ω, f is a known data function, g is a known boundary function, and ε and $\boldsymbol{b}$ are coefficients for diffusion and convection, respectively. For $\varepsilon \ll |\boldsymbol{b}|$ we say that this represents a convection-dominated diffusion problem. In such cases, solutions tend to develop boundary layers, that is, components of the solution that have steep gradients near the boundary. To illustrate this, consider the following ordinary differential equation analogy:

$$\begin{cases} -\varepsilon u'' + bu' = 0 & \text{in } (0, 1), \\ u(0) = 1, \\ u(1) = 0, \end{cases} \tag{2}$$

MSC2010: 65N30, 65N50, 35J57.

Keywords: partial differential equations, finite element methods, convection, boundary layers.

This work was sponsored by the National Science Foundation under grant DMS-0755260 through the Wabash Summer Institute in Mathematics at Wabash College.

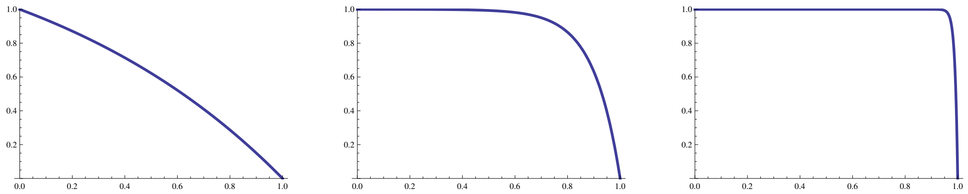


Figure 1. Solution of (2) for $\varepsilon = 1$ (left), $\varepsilon = 0.1$ (middle), and $\varepsilon = 0.01$ (right).

where bu' is the convection term and $-\varepsilon u''$ the diffusion term. We call the ODE convection-dominated when $\varepsilon \ll |b|$, and to illustrate this we set $b = 1$ and consider the following solution plots for different values of ε in Figure 1.

We can see that as $\varepsilon \to 0$, a boundary layer forms near $x = 1$. This behavior is difficult to approximate computationally and is also present in the solution of system (1) for regions of Ω near boundary points with $\boldsymbol{n} \cdot \boldsymbol{b} > 0$, where $\boldsymbol{n}$ is an outward unit normal to $\partial\Omega$. See [Brenner and Scott 1994; Braess 2001] for background on finite element methods for such problems.

The method we develop here is a generalization of a least squares finite element discretization for scalar elliptic equations. In general, a least squares approach to (1) tends to be an effective way to approximate solutions; however, convergence is degraded in the presence of dominant convection. We consider a least squares functional minimized with respect to a weighted L^2-norm, where the weights are chosen adaptively in the context of an adaptive mesh refinement routine. This idea is inspired by work of Westphal et al. [Lee et al. 2006; 2008; Cai and Westphal 2008], where a weighted functional is used to improve solutions to problems with singularities.

The organization of this paper is as follows: in Section 2 we introduce a reformulated version of (1) and the adaptively weighted procedure; in Section 3 we provide several numerical tests to demonstrate the effectiveness of the method compared to a more standard approach; and in Section 4 we show the robustness of the idea by applying it analogously to a moderately high Reynolds number Navier–Stokes system for steady fluid flow.

2. Methodology

The $L^2(\Omega)$ norm of a function f is defined to be

$$\|f\| = \|f\|_{L^2(\Omega)} = \left(\int_\Omega |f|^2\right)^{1/2},$$

and $L^2(\Omega)$ is the space of functions in Ω that have finite $L^2(\Omega)$ norms. Likewise, we define $H^1(\Omega)$ as the subspace of $L^2(\Omega)$ where all first partial derivatives of functions are also in $L^2(\Omega)$.

With the substitution $\boldsymbol{\sigma} = -\varepsilon\nabla u$, we rewrite (1) as the first-order equations

$$\begin{cases} \nabla\cdot\boldsymbol{\sigma} + \boldsymbol{b}\cdot\nabla u = 0 & \text{in } \Omega, \\ \boldsymbol{\sigma} + \varepsilon\nabla u = \boldsymbol{0} & \text{in } \Omega, \\ \nabla\times\boldsymbol{\sigma} = 0 & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \\ \hat{\tau}\cdot\boldsymbol{\sigma} = -\varepsilon\hat{\tau}\cdot\nabla g & \text{on } \partial\Omega. \end{cases} \tag{3}$$

The third equation holds because $\nabla\times\boldsymbol{\sigma} = \nabla\times(-\varepsilon\nabla u) = -\varepsilon(\nabla\times\nabla u) = 0$. In the fifth equation, $\hat{\tau}$ is a unit tangent vector to $\partial\Omega$ and this new boundary condition is simply a statement about the directional derivative of g along $\partial\Omega$.

We first consider what we refer to as the standard least squares approach. Since we seek a finite element solution, we partition Ω into an initial triangulation denoted as Ω^h. Here, h denotes the size, or width of the triangles and $(u^h, \boldsymbol{\sigma}^h)$ represents an approximate solution to $(u, \boldsymbol{\sigma})$, the exact solution of system (3). Define

$$\begin{aligned} \boldsymbol{V} &= \{v \in H^1(\Omega) : v = g \text{ on } \partial\Omega\}, \\ \boldsymbol{\Sigma} &= \{\boldsymbol{s} \in H^1(\Omega)^2 : \tau\cdot\boldsymbol{s} = -\varepsilon\tau\cdot\nabla g \text{ on } \partial\Omega\} \end{aligned}$$

as sets of admissible solutions, and let $\boldsymbol{V}^h \subseteq \boldsymbol{V}$ and $\boldsymbol{\Sigma}^h \subseteq \boldsymbol{\Sigma}$ be finite dimensional subsets in which we seek approximate solutions.

A standard LS approach seeks a pair of solutions $(u^h, \boldsymbol{\sigma}^h) \in \boldsymbol{V}^h \times \boldsymbol{\Sigma}^h$ which minimizes the functional

$$G(u^h, \boldsymbol{\sigma}^h; f) = \|\nabla\cdot\boldsymbol{\sigma}^h + \boldsymbol{b}\cdot\nabla u^h - f\|^2 + \|\boldsymbol{\sigma}^h + \varepsilon\nabla u^h\|^2 + \|\nabla\times\boldsymbol{\sigma}^h\|^2. \tag{4}$$

For elliptic problems that are diffusion dominated, minimizing (4) using standard finite element spaces results in good convergence. However, for convection-dominated problems, minimizing (4) results in slow convergence until h is very small (typically $h \approx \mathbb{O}(\varepsilon)$). Other finite element approaches tend to be unstable in convection-dominated regimes and solutions may exhibit oscillatory behavior; see, e.g., [Bochev and Gunzburger 2009; Strang and Fix 1973].

One undesirable aspect of the standard least squares approach is that there is not only significant error near boundary layers, but that the error may remain large even in regions of the domain where the solution is smooth. To reduce this "pollution effect", we introduce weight functions into the functional (4) to redefine the metric of the approximation space. By doing this, we are able to force the least squares functional to choose a better solution globally (i.e., in the regions of the domain where the solution is smooth) and segregate errors to a small region near boundary layers. Thus, we want to choose the weight function, w, to be large (a value at or near 1) where the error is small, and small (a value near 0) where the error is large.

In this paper we use what we call a sigma-based weighting strategy that uses the approximate solution for $\boldsymbol{\sigma}$ to construct the weight function. An alternative, one we refer to as functional based weighting, uses locally evaluated functional values to generate weights. Though both strategies have merits, we focus here on sigma-based weights. Consider an approximate solution $\boldsymbol{\sigma}^h$ evaluated on a single finite element triangle, T:

$$\|\boldsymbol{\sigma}^h\|_T = \left(\int_T |\boldsymbol{\sigma}^h|^2\right)^{1/2},$$

which we may use as a local indicator of where the solution is likely to have steep gradients (recall the definition of $\boldsymbol{\sigma}$). We thus choose a weight function, w, on each T by the procedure illustrated in Figure 2.

We choose $w_{min} = e^{-h/\varepsilon}$. For coarse meshes, where weighting is most needed, w_{min} is very near zero. For increasingly fine meshes, where the weight procedure is needed less, we have $\lim_{h\to 0} w_{min} \to 1$. Thus our algorithm remains robust for a wide range of convection-diffusion regimes.

With such an appropriate weight function chosen we find an improved approximate solution by choosing u^h and $\boldsymbol{\sigma}^h$ that minimize the weighted least squares functional

$$G(u^h, \boldsymbol{\sigma}^h; f) \\ = \|w(\nabla\cdot\boldsymbol{\sigma}^h + \boldsymbol{b}\cdot\nabla u^h - f)\|^2 + \|w(\boldsymbol{\sigma}^h + \varepsilon\nabla u^h)\|^2 + \|w(\nabla\times\boldsymbol{\sigma}^h)\|^2, \quad (5)$$

where we note that setting $w = 1$ corresponds to the original least squares functional (4). Since this approach obviously requires an initial approximate solution to choose w, it makes sense to conduct this in a nested iteration approach where the initial approximation is found cheaply on a coarse mesh and the improved approximation is found on refined mesh. In other words, our approach is to incorporate

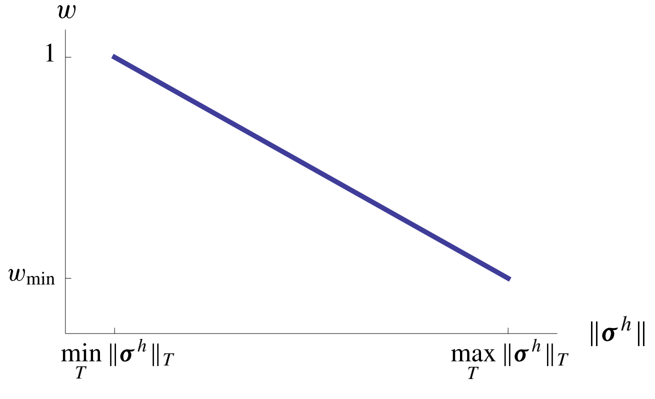


Figure 2. The relationship between $\|\boldsymbol{\sigma}^h\|_T$ and the weight function.

refining the weight function in (5) into an adaptive mesh refinement routine for finding increasingly accurate approximations on a sequence of refined meshes.

We describe the solution process in the following algorithm:

- **Start:** Consider minimizing (5) on an initial coarse triangulation Ω^H, where H is the mesh size. Initially set $w = 1$.
- **Coarse solve:** Minimize (5) to find $(u^H, \boldsymbol{\sigma}^H) \in \boldsymbol{V}^H \times \boldsymbol{\Sigma}^H$.
- **Construct weights:** Using the rule illustrated in Figure 2, choose w to be a piecewise linear function on each element in Ω^H.
- **Refine mesh:** The locally evaluated least squares functional is used to determine triangles in Ω^H with the highest concentration of error, which are refined by splitting each into four smaller triangles. Let $h = H/2$ represent the mesh size of the refined mesh, Ω^h.
- **Fine solve:** Minimize (5) to find $(u^h, \boldsymbol{\sigma}^h) \in \boldsymbol{V}^h \times \boldsymbol{\Sigma}^h$. Set $H \leftarrow h$ as the coarse scale for the problem and repeat the procedure.

Figure 3 illustrates the multilevel iterative algorithm.

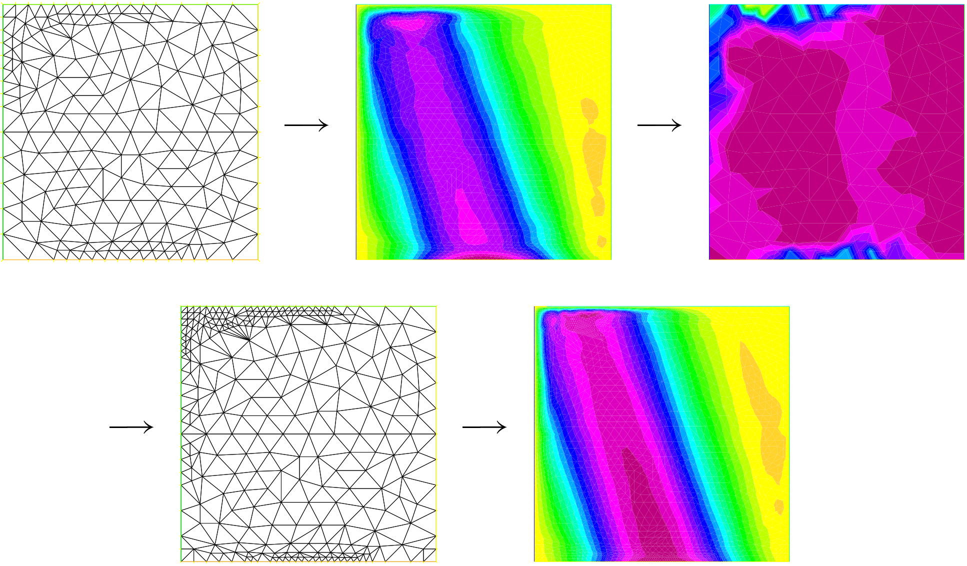


Figure 3. Iterative process for computing approximate solutions: coarse mesh, coarse solution, weight function, refined mesh, fine solution.

3. Testing and results

We test several problems with various levels of difficulty. We compare our approximate solution $(u^h, \boldsymbol{\sigma}^h)$ to a control solution to get the associated error. This control solution is obtained by computing the solution on a superfine scale mesh using the standard LS approach over several iterations. We assume it to be sufficiently accurate for our purpose of comparison. We compute the L^2 norm of this error as a measure of accuracy of the approximated solution. In all cases we use conforming piecewise quadratic finite elements for each unknown. In the computational tests in this section, we choose $\Omega = (0, 1)^2$ and zero Dirichlet boundary conditions on the north, east and west boundaries, and define a nonzero $g(x)$ on the south boundary.

The following four examples compare the efficiency of the standard LS approach and our sigma-based weighting strategy. Both axes in all graphs are on a $\log_{10}$ scale. The points that are higher have larger errors than the lower ones.

Example 1. We solve the system (1) with a constant $\boldsymbol{b} = \left(-\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$, a smooth $g = 16x^2(1-x)^2$, and a relatively large $\varepsilon = 0.005$. The results are shown in Figure 4; it can be seen that our sigma-based weighting method yields a more accurate solution (by a factor of 3 approximately) than the standard LS approach.

Example 2. Next we take a nonconstant convection coefficient,

$$\boldsymbol{b} = \left(\frac{-y}{\sqrt{x^2+y^2}}, \frac{x}{\sqrt{x^2+y^2}}\right),$$

with g and ε as in Example 1. The results, shown in Figure 5, show that our approach still outperforms the standard one, though by a lesser factor than before.

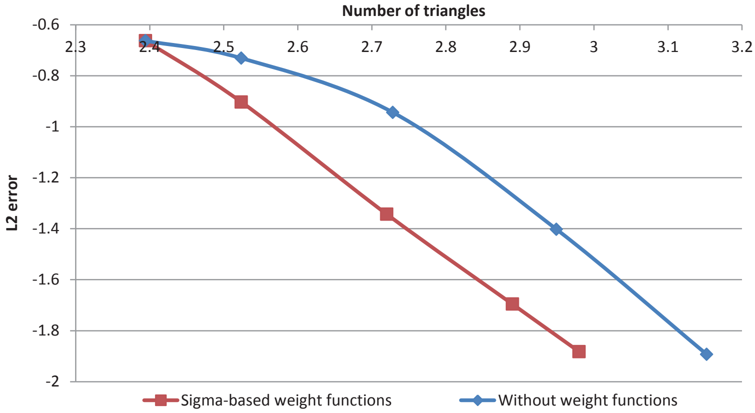


Figure 4. Log-log plot of L^2 norm of error (with respect to control solution) as a function of the number of triangles, for $\boldsymbol{b} = \left(-\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$, $g = 16x^2(1-x)^2$, $\varepsilon = 0.005$ (Example 1).

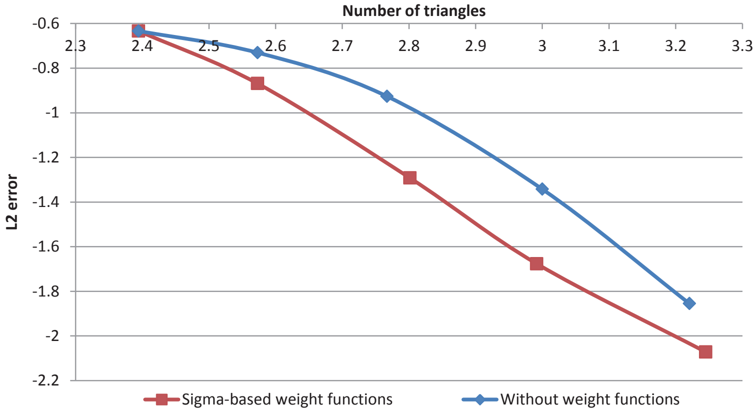


Figure 5. Like Figure 4, with $\boldsymbol{b} = \left(-y/\sqrt{x^2+y^2}, x/\sqrt{x^2+y^2}\right)$, $g = 16x^2(1-x)^2$, $\varepsilon = 0.005$ (Example 2).

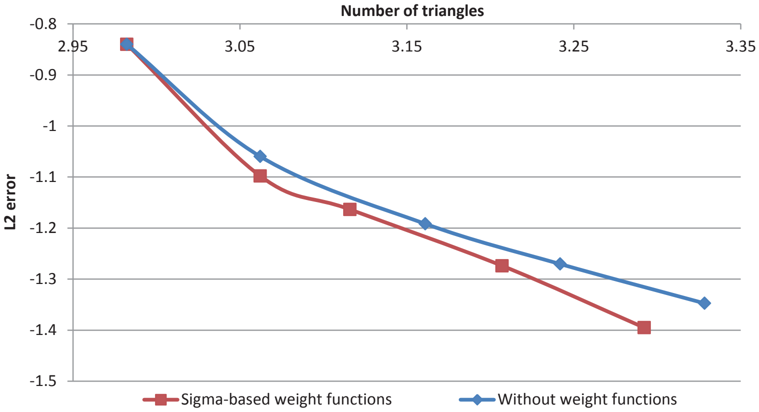


Figure 6. Like Figure 4, with g discontinuous (Example 3).

Example 3. We return to $\boldsymbol{b}$ and ε as in Example 1, and choose a discontinuous boundary function,

$$g = \begin{cases} 1 & \text{if } x \in (0.2, 0.8), \\ 0 & \text{else}. \end{cases}$$

Here the two curves (Figure 6) come even closer than in the previous example, but the solution with sigma-based weights is still the more accurate one. With discontinuous data on the boundary, the solution here is much more difficult to approximate numerically, so the overall error is larger than the previous examples.

Example 4. For our final example in this section, we decrease ε by an order of magnitude, that is, $\varepsilon = 0.0005$, while $\boldsymbol{b} = \left(-\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$ and $g = 16x^2(1-x)^2$ stay

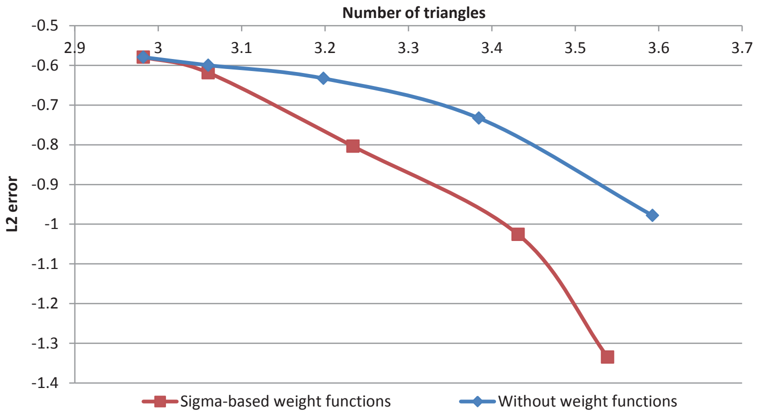


Figure 7. Like Figure 4, with $\boldsymbol{b} = \big(-\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\big)$, $g = 16x^2(1-x)^2$, $\varepsilon = 0.0005$.

the same as in Example 1. Here again, our method shows an improvement over the standard approach (Figure 7).

4. Results for Navier–Stokes equations

The preceding examples suggest that the sigma-based weighting method is generally more efficient than the standard LS approach. As a further case study, we consider a more complicated system of equations that retains the same set of challenges as the convection-dominated diffusion problem.

In this section we directly apply the adaptively weighted norm minimization strategy to a more difficult system of equations. We consider the stationary incompressible Navier–Stokes equations in the form

$$\begin{cases} -\dfrac{1}{\mathrm{Re}}\Delta \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p = \boldsymbol{f} & \text{in } \Omega, \\ \nabla\cdot\boldsymbol{u} = 0 & \text{in } \Omega, \\ \boldsymbol{u} = \boldsymbol{g} & \text{on } \partial\Omega, \end{cases} \tag{6}$$

where $\boldsymbol{u}$ denotes the velocity of fluid flow in the x and y direction, p the pressure of fluid flow, $\boldsymbol{f}$ a given body force, and Re denotes the Reynolds number, a measure of the potential turbulence of the fluid. With the two substitutions

$$\varepsilon = \frac{1}{\mathrm{Re}} \quad \text{and} \quad \boldsymbol{U} = -\varepsilon\nabla\boldsymbol{u},$$

the first equation in (6) becomes

$$\nabla\cdot\boldsymbol{U} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p = \boldsymbol{f}. \tag{7}$$

Utilizing a Newton linearization, we have the following approximation

$$\boldsymbol{u} \cdot \nabla \boldsymbol{u} \approx \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}_{\text{old}} - \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u}_{\text{old}},$$

where $\boldsymbol{u}_{\text{old}} \approx \boldsymbol{u}$ is a known approximation to $\boldsymbol{u}$. This current solution, $\boldsymbol{u}_{\text{old}}$, is initially set to be $(0, 0)$. During the iteration process, each time we obtain a new approximate solution to $\boldsymbol{u}$, we assign its value to $\boldsymbol{u}_{\text{old}}$. Therefore, the older $\boldsymbol{u}_{\text{old}}$ in (6) will be replaced by the new one to better approximate the left-hand side of (6). After the substitution, (6) is reformulated as

$$\begin{cases} \nabla \cdot \boldsymbol{U} + \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}_{\text{old}} + \nabla p = \boldsymbol{f} + \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u}_{\text{old}} & \text{in } \Omega, \\ \boldsymbol{U} + \varepsilon \nabla \boldsymbol{u} = \boldsymbol{0} & \text{in } \Omega, \\ \nabla \times \boldsymbol{U} = \boldsymbol{0} & \text{in } \Omega, \\ \nabla \cdot \boldsymbol{u} = 0 & \text{in } \Omega, \\ \boldsymbol{u} = \boldsymbol{g} & \text{on } \partial\Omega. \end{cases}$$

Notice the similarity between this system and (3), which gives us confidence that the weighted norm procedure can improve a least squares solution method for this system. For large Re, turbulent flow characteristics, including boundary layers, may develop, which is similar to the behavior of convection-dominated PDEs. Therefore, we define our weighted, linearized LS functional to be

$$\begin{aligned} G(\mathbf{u}, \mathrm{U}, p; \boldsymbol{f}) = \|w(\nabla \cdot \boldsymbol{U} + \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}_{\text{old}} + \nabla p) - (\boldsymbol{f} + \boldsymbol{u}_{\text{old}} \cdot \nabla \boldsymbol{u}_{\text{old}})\|^2 \\ + \|w(\boldsymbol{U} + \varepsilon \nabla \boldsymbol{u})\|^2 + \|w(\nabla \times \boldsymbol{U})\|^2 + \|w(\nabla \cdot \boldsymbol{u})\|^2, \end{aligned}$$

where w denotes our weight function. On each mesh we carry out several steps of Newton linearization, and then adaptively refine our mesh. Weight functions are now constructed based on $\boldsymbol{U}$ (which is analogous to $\boldsymbol{\sigma}$ in the convection-dominated diffusion system).

To test our weighting strategy, we choose our domain Ω to be $(0, 1)^2 \setminus (0, 0.5]^2$. We set $\varepsilon = 1/200$ and $\boldsymbol{u} = ((1 - e^{-(y-0.5)/\varepsilon})(1 - e^{-(1-y)/\varepsilon}), 0)$ on the upper west boundary and $\boldsymbol{u} = (0, -(1 - e^{-(x-0.5)/\varepsilon})(1 - e^{-(1-x)/\varepsilon}))$ on the south boundary. We again set $\boldsymbol{f}$ to be 0 for simplicity. Figure 8 shows the control solution for our test problem.

We set the number of Newton linearization steps on each mesh to 3. Figure 9 compares the accuracy of both approaches to solving Navier–Stokes equations. The x- and y-axis are on a $\log_{10}$ scale.

The result shows again that our sigma-based weighting method is more efficient at solving the Navier–Stokes equations than having no weight functions.

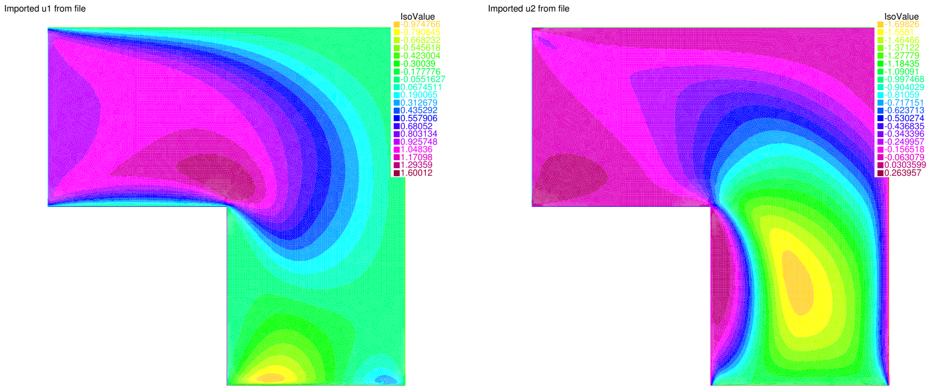


Figure 8. The control solution for u_1 (left) and u_2 (right), obtained on a fine mesh and presumed very accurate.

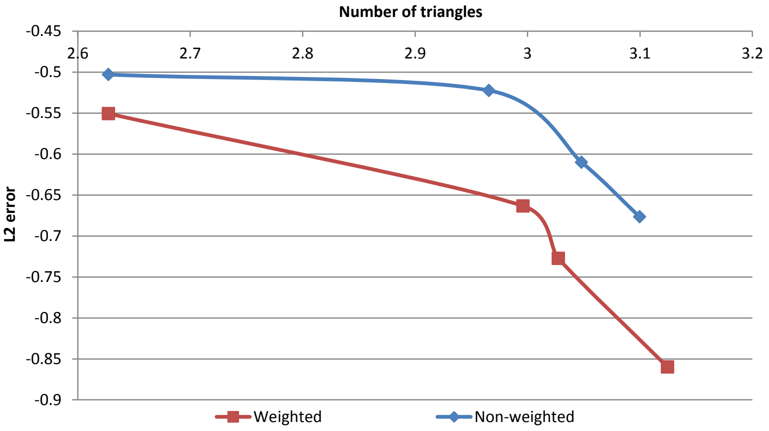


Figure 9. Comparison of weighted and nonweighted approaches for the Navier–Stokes example (log-log plot).

5. Conclusion

We find that defining and adaptively modifying weight functions in a least squares functional can improve the efficiency of the method for convection-dominated problems. Our approach uses approximate solutions on coarse meshes to adapt the metric of the approximation space so that the error is reduced with respect to a better scaled norm than a standard approach. The procedure is easily adapted to more difficult convection-dominated problems, such as the steady Navier–Stokes equations.

References

[Bochev and Gunzburger 2009] P. B. Bochev and M. D. Gunzburger, *Least-squares finite element methods*, Applied Math. Sci. **166**, Springer, New York, 2009. MR 2010b:65004 Zbl 1168.65067

[Braess 2001] D. Braess, *Finite elements: theory, fast solvers, and applications in solid mechanics*, 2nd ed., Cambridge University Press, Cambridge, 2001. MR 2001k:65002 Zbl 0976.65099

[Brenner and Scott 1994] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, Texts in Applied Mathematics **15**, Springer, New York, 1994. MR 95f:65001 Zbl 0804.65101

[Cai and Westphal 2008] Z. Cai and C. R. Westphal, "A weighted $H(\text{div})$ least-squares method for second-order elliptic problems", *SIAM J. Numer. Anal.* **46**:3 (2008), 1640–1651. MR 2008m:65314 Zbl 1168.65069

[Lee et al. 2006] E. Lee, T. A. Manteuffel, and C. R. Westphal, "Weighted-norm first-order system least squares (FOSLS) for problems with corner singularities", *SIAM J. Numer. Anal.* **44**:5 (2006), 1974–1996. MR 2008a:65221 Zbl 1129.65087

[Lee et al. 2008] E. Lee, T. A. Manteuffel, and C. R. Westphal, "Weighted-norm first-order system least-squares (FOSLS) for div/curl systems with three dimensional edge singularities", *SIAM J. Numer. Anal.* **46**:3 (2008), 1619–1639. MR 2009c:65316 Zbl 1170.65095

[Strang and Fix 1973] G. Strang and G. J. Fix, *An analysis of the finite element method*, Prentice-Hall, Englewood Cliffs, NJ, 1973. MR 56 #1747 Zbl 0356.65096

Received: 2010-12-20 Accepted: 2011-09-04

bkraynik11@alumnimail.wooster.edu
Department of Mathematics and Computer Science,
College of Wooster, Wooster, OH 44691, United States

ysun13@wabash.edu
Department of Mathematics and Computer Science,
Wabash College, Crawfordsville, IN 47933, United States

westphac@wabash.edu
Department of Mathematics and Computer Science,
Wabash College, Crawfordsville, IN 47933, United States

INVOLVE 5:1(2012)

Diameter, girth and cut vertices of the graph of equivalence classes of zero-divisors

Blake Allen, Erin Martin, Eric New and Dane Skabelund

(Communicated by Scott Chapman)

We explore the properties of $\Gamma_E(R)$, the graph of equivalence classes of zero-divisors of a commutative Noetherian ring R. We determine the possible combinations of diameter and girth for the zero-divisor graph $\Gamma(R)$ and the equivalence class graph $\Gamma_E(R)$, and examine properties of cut-vertices of $\Gamma_E(R)$.

Introduction

The zero-divisor graph of a commutative ring R, was first introduced in [Beck 1988] and has since been investigated in various forms. It was shown in [Anderson and Livingston 1999] that the zero-divisor graph of any ring is connected with diameter less than or equal to 3. Mulay [2002] proved many interesting results about cycles in the zero-divisor graph.

In 2009, Spiroff and Wickham [2011] introduced $\Gamma_E(R)$, the graph of equivalence classes of zero-divisors, which is a simplification of the zero-divisor graph $\Gamma(R)$. The vertices of $\Gamma_E(R)$ are, instead of individual zero-divisors of R, equivalence classes of zero-divisors determined by annihilator ideals. The graph $\Gamma_E(R)$ provides a more succinct view of the zero-divisor activity of the ring. In many cases, the equivalence class graph is finite even though the zero-divisor graph is infinite. For example, for $S = \mathbb{Z}[X, Y]/(X^4, XY)$, the graph $\Gamma(S)$ is infinite, while the graph $\Gamma_E(S)$ has only 6 vertices. Specifically, the vertices corresponding to $X^3, 2X^3, 3X^3, \ldots$ are all distinct in $\Gamma(S)$. However, since they all have the same annihilator, they all belong to the same equivalence class, and so are represented by a single vertex $[X^3]$ in $\Gamma_E(S)$.

The equivalence class graph also lets us view the interplay between the annihilator ideals of R and helps to easily identify the associated primes of the ring. The vertices of $\Gamma_E(R)$ which correspond to associated primes have special properties which will help us to prove several interesting results related to $\Gamma_E(R)$. In

MSC2010: 13A99.

Keywords: zero-divisor graph, diameter, girth, cut vertices.

Section 1, we provide basic definitions and background. In Section 2, we determine all possible diameter combinations of $\Gamma(R)$ and $\Gamma_E(R)$, and do the same for the girth of the two graphs in Section 3. In Section 4, we look at properties of the cut-vertices of $\Gamma_E(R)$. Throughout, R will denote a commutative Noetherian ring.

1. Background and basic results

Graph theory. We briefly review basic graph theory terms that we will use throughout the paper. All graphs we deal with will be *simple* graphs in the sense that they contain no loops or double edges. We will denote the set of vertices of a graph Γ by $V(\Gamma)$. If two vertices x and y are joined by an edge, we say x and y are *adjacent*, and write $x - y$. A *path* is defined as an alternating sequence of distinct vertices and edges, and the *length of a path* is the number of edges in the path. If x and y are two vertices, then the *distance between x and y*, denoted $d(x, y)$, is the length of the shortest path from x to y. If there is no path connecting x to y, we say that $d(x, y) = \infty$, and we define $d(x, x) = 0$. The *diameter of a graph* is the maximum distance between any two vertices of the graph. We will denote the diameter of a graph Γ by $\operatorname{diam}\Gamma$. A *cycle* is a closed path, or a path that starts and ends on the same vertex. The *girth* of a graph is the length of its smallest cycle. We denote the girth of a graph Γ by $\mathrm{g}(\Gamma)$ and say that $\mathrm{g}(\Gamma) = \infty$ if the graph Γ contains no cycle. Note that the smallest possible cycle length is 3, so if Γ contains a cycle, $\mathrm{g}(\Gamma) \geq 3$.

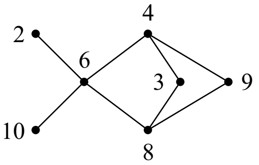
A graph is said to be *connected* if every pair of vertices is joined by a path and *complete* if every pair of vertices is joined by an edge. A *connected component* of a graph Γ is a maximal connected subgraph of Γ. If removing a vertex v from a graph along with all its incident edges increases the number of connected components in the graph, then v is called a *cut vertex*. A graph is *complete bipartite* if its vertices can be partitioned into two subsets, V_1 and V_2, such that every vertex of V_1 is adjacent to every vertex of V_2, but no two vertices of V_1 are adjacent and no two vertices of V_2 are adjacent. Such a graph will be denoted $K_{n,m}$, where $n = |V_1|$ and $m = |V_2|$. If the vertices of a graph can be partitioned into r subsets in a similar fashion, then the graph is said to be *r-partite*.

Zero-divisor graphs. Let $Z(R)$ denote the set of zero-divisors of R and $Z^*(R)$ denote the set $Z(R) \setminus \{0\}$. We define the *zero-divisor graph* of R as the simple graph $\Gamma(R)$ where the vertices of $\Gamma(R)$ are the elements of $Z^*(R)$, and there is an edge between $x, y \in \Gamma(R)$ whenever $xy = 0$.

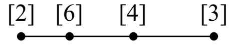
Recall that the annihilator ideal associated to an element $x \in R$ is the set $\operatorname{ann} x = \{r \in R : xr = 0\}$. We define an equivalence relation $\sim$ on R such that for all $x, y \in R$, we say $x \sim y$ if $\operatorname{ann} x = \operatorname{ann} y$. Let $[x]$ denote the equivalence class of x. Notice that $[0] = \{0\}$, $[1] = R \setminus Z(R)$ and the relation $\sim$ partitions the remaining zero-divisors into distinct classes. Furthermore, it follows that the multiplication of these equivalence classes $[x] \cdot [y] = [xy]$ is well-defined.

The *graph of equivalence classes of zero-divisors of* R, $\Gamma_E(R)$, is the graph whose vertices are the classes of nonzero zero-divisors of R determined by the relation $\sim$, where there is an edge between two vertices $[x]$ and $[y]$ if $[x]\cdot[y] = [0]$.

Here, as an example, are the zero-divisor graph of $\mathbb{Z}_{12}$ and the graph of its equivalence classes:



$\Gamma(\mathbb{Z}_{12})$



$\Gamma_E(\mathbb{Z}_{12})$

We see that since $\operatorname{ann} 2 = \operatorname{ann} 10$, the elements 2 and 10 are in the same equivalence class, and therefore collapse to the single vertex [2] in $\Gamma_E(R)$.

Previous results. Spiroff and Wickham [2011] have several interesting results linking the associated primes of R with the structure of $\Gamma_E(R)$. These will be useful in furthering our investigation of $\Gamma_E(R)$. Remember that a prime ideal $\mathfrak{p}$ of R is an *associated prime* if it is the annihilator of some element of R. The set of associated primes is denoted $\operatorname{ass} R$. It is well known that if R is a Noetherian ring, then $\operatorname{ass} R$ is nonempty and finite and that any maximal element of the family of annihilator ideals $\mathfrak{F} = \{\operatorname{ann} x : 0 \neq x \in R\}$ is an associated prime. Note also that since every zero divisor is contained in an annihilator ideal and maximal annihilators are associated primes, the set of zero-divisors of R equals the union of all associated primes of R. Since there is exactly one vertex of $\Gamma_E(R)$ for each distinct annihilator ideal of R, we have a natural injection of $\operatorname{ass} R$ into the vertex set of $\Gamma_E(R)$ given by $\mathfrak{p} \mapsto [y]$ where $\mathfrak{p} = \operatorname{ann} y$. We adopt the conventions of Spiroff and Wickham and by a slight abuse of terminology will refer to the vertex $[y]$ as an associated prime if $\operatorname{ann} y \in \operatorname{ass} R$. It will be clear from context whether $[y]$ refers to an equivalence class, a vertex, or a specific annihilator.

Lemma 1.1 [Spiroff and Wickham 2011, Lemma 1.2]. *Any two distinct elements of* $\operatorname{ass} R$ *are connected by an edge. Furthermore, every vertex* $[v]$ *of* $\Gamma_E(R)$ *is either an associated prime or adjacent to an associated prime maximal in* $\mathfrak{F}$.

Lemma 1.2 [Spiroff and Wickham 2011, Proposition 1.7]. *Let R be a ring such that $\Gamma_E(R)$ is complete r-partite. Then $r = 2$ and $\Gamma_E(R) = K_{n,1}$ for some $n \geq 1$.*

2. Diameter

In this section, we explore the relationship between the diameters of the graphs $\Gamma(R)$ and $\Gamma_E(R)$. It is shown in [Anderson and Livingston 1999] that $\Gamma(R)$ has diameter at most 3 for any commutative ring R. In [Spiroff and Wickham 2011] it is shown that $\operatorname{diam}\Gamma_E(R) \leq 3$ for R commutative and Noetherian. The following results further demonstrate the relationship between the diameters of the two graphs.

Proposition 2.1. *If R is a commutative ring, then* $\operatorname{diam}\Gamma_E(R) \leq \operatorname{diam}\Gamma(R)$.

Proof. Let $[a], [b] \in \Gamma_E(R)$ with $d([a], [b]) = n$, and let $[a] = [x_1] - [x_2] - \cdots - [x_{n+1}] = [b]$ be a path of minimal length from $[a]$ to $[b]$. From each $[x_i]$, choose one $y_i \in [x_i]$. Then $y_1 - y_2 - \cdots - y_{n+1}$ is a path in $\Gamma(R)$ of length n. We claim that this path is minimal, and thus $d(y_1, y_{n+1}) = n$. If this path is not minimal, there is some shorter path $y_1 = z_1 - z_2 - \cdots - z_{m+1} = y_{n+1}$, with $m < n$. Since either $[z_i] = [z_{i+1}]$ or $[z_i] - [z_{i+1}]$, the path $[y_1] = [z_1] - [z_2] - \cdots - [z_{m+1}] = [y_{n+1}]$ has length less than or equal to m, a contradiction. □

Theorem 2.2. *If* $\operatorname{diam}\Gamma_E(R) = 0$, *then* $\operatorname{diam}\Gamma(R) = 0$ *or* 1.

Proof. Let $\Gamma_E(R)$ have diameter 0. Since $\Gamma_E(R)$ has only one vertex, $[x] = [y]$ for every $x, y \in Z^*(R)$. Since the graph $\Gamma(R)$ is connected and every element in $\Gamma(R)$ has the same annihilator, $xy = 0$ for every $x, y \in Z^*(R)$. Thus the graph $\Gamma(R)$ is complete and $\operatorname{diam}\Gamma(R) = 0$ or 1. □

Theorem 2.3. *If* $\operatorname{diam}\Gamma(R) = 3$, *then* $\operatorname{diam}\Gamma_E(R) = 3$.

Proof. Let $\Gamma(R)$ have diameter 3. Then for some elements $x, w \in \Gamma(R)$, $d(x, w) = 3$ in $\Gamma(R)$. Let $x - y - z - w$ be a path from x to w of minimal length. Since this path is minimal, $xz \neq 0$, but $zw = 0$, so $\operatorname{ann} x \neq \operatorname{ann} w$. By similar reasoning we see that each of $\operatorname{ann} x$, $\operatorname{ann} y$, $\operatorname{ann} z$, and $\operatorname{ann} w$ are distinct. Hence $[x]$, $[y]$, $[z]$, and $[w]$ are distinct equivalence classes in $\Gamma_E(R)$. Thus $[x]$ is not adjacent to $[w]$ and there exist no paths $[x]-[y]-[w]$ or $[x]-[z]-[w]$ in $\Gamma_E(R)$. Now suppose there is some other $[v]$ such that $[x]-[v]-[w]$. This is impossible because it implies that there is a path $x-v-w$ in $\Gamma(R)$, contradicting the supposition that $x-y-z-w$ is a minimal path. Therefore $d([x], [w]) = 3$ and since $\operatorname{diam}\Gamma_E(R) \leq 3$, $\operatorname{diam}\Gamma_E(R) = 3$. □

We summarize with Table 1, which shows all possible combinations of diameter for $\Gamma(R)$ and $\Gamma_E(R)$.

We see from our examples that it is possible for the diameter of the zero-divisor graph to shrink under the equivalence relation. We consider the situations where this happens.

If $\operatorname{diam}\Gamma(R) = 1$ and $\operatorname{diam}\Gamma_E(R) = 0$, then R has a unique annihilator ideal $\operatorname{ann} x$. This annihilator is maximal in $\mathfrak{F}$ and an associated prime of the ring. Since $Z(R) = \bigcup_{\mathfrak{p}\in\operatorname{ass} R} \mathfrak{p} = \operatorname{ann} x$, $Z(R)$ forms an ideal of R.

diam $\Gamma(R)$	diam $\Gamma_E(R) =$ 0	1	2	3
0	$\mathbb{Z}_4$, $\mathbb{Z}_2[x]/(x^2)$	–	–	–
1	$\mathbb{Z}_9$, $\mathbb{Z}_{25}$	$\mathbb{Z}_2 \times \mathbb{Z}_2$	–	–
2	impossible (Theorem 2.2)	$\mathbb{Z}_6$, $\mathbb{Z}_{21}$, $\mathbb{Z}_2[x]/(x^3)$	$\mathbb{Z}_{16}$	–
3	impossible (Theorem 2.3)			$\mathbb{Z}_{12}$

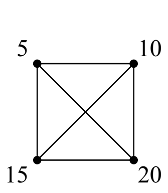
Table 1. Possibilities for diam $\Gamma(R)$ and diam $\Gamma_E(R)$, with examples.

Next we consider the situation in which the diameter reduces from 2 to 1. Since there are no complete equivalence class graphs on 3 or more vertices, by [Spiroff and Wickham 2011, Proposition 1.5], $\Gamma_E(R)$ must have exactly two vertices, and R must have exactly 2 distinct annihilator ideals, $\operatorname{ann} x$ and $\operatorname{ann} y$. Let $\operatorname{ann} x$ be maximal in $\mathfrak{F}$. If $\operatorname{ann} y \subseteq \operatorname{ann} x$, then $Z(R) = \bigcup_{\mathfrak{p} \in \operatorname{ass} R} \mathfrak{p} = \operatorname{ann} x$ forms an ideal of R. Otherwise, both $\operatorname{ann} x$ and $\operatorname{ann} y$ are maximal in $\mathfrak{F}$ and $\operatorname{ann} x \cap \operatorname{ann} y = \{0\}$. If we have nonzero a, b with $a \in \operatorname{ann} x$ and $b \in \operatorname{ann} y$ such that $a + b \in \operatorname{ann} x$, then $b \in \operatorname{ann} x$, a contradiction. So in this case $Z(R) = \bigcup_{\mathfrak{p} \in \operatorname{ass} R} \mathfrak{p} = \operatorname{ann} x \cup \operatorname{ann} y$ does not form an ideal of R.

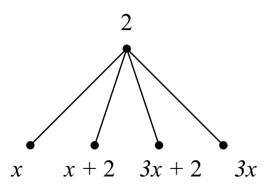
Therefore we see that if the diameter shrinks in the equivalence class graph, R has 1 or 2 associated primes. If R is a finite ring, this corresponds to R being the direct product of 1 or 2 local rings, since every finite ring R is expressible as the product of finite local rings, with the number of factors equal to the number of associated primes of R.

We show below examples of graphs of rings with shrinking diameter, one from each of the situations considered above. Note that $\mathbb{Z}_{25}$ has a unique annihilator,

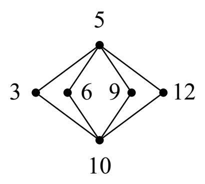
$\Gamma(\mathbb{Z}_{25})$



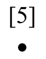
$\Gamma(\mathbb{Z}_4[x] / (2x, x^2-2))$



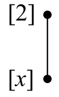
$\Gamma(\mathbb{Z}_{15})$



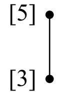
$\Gamma_E(\mathbb{Z}_{25})$



$\Gamma_E(\mathbb{Z}_4[x] / (2x, x^2-2))$



$\Gamma_E(\mathbb{Z}_{15})$



$\operatorname{ann} 5 = (5)$, $\mathbb{Z}_4[x]/(2x, x^2 - 2)$ has two annihilators, $\operatorname{ann} x = (2) \subsetneq (2, x) = \operatorname{ann} 2$, and $\mathbb{Z}_{15}$ has two annihilators, $\operatorname{ann} 3 = (5)$ and $\operatorname{ann} 5 = (3)$, which intersect trivially.

3. Girth

Mulay [2002] proved that if the zero-divisor graph, $\Gamma(R)$, contains a cycle then $\mathrm{g}(\Gamma(R)) \leq 4$. In this section we will demonstrate an even stronger restriction on the girth of the equivalence class graph, and find all possible combinations of girth for $\Gamma(R)$ and $\Gamma_E(R)$. The following result gives a girth restriction for $\Gamma_E(R)$ similar to that shown by Mulay for $\Gamma(R)$.

Theorem 3.1. *If R is a commutative Noetherian ring, and if $\Gamma_E(R)$ contains a cycle, then $\mathrm{g}(\Gamma_E(R)) \leq 4$.*

Proof.

Case 1: If R has at least 3 distinct associated primes, say $\operatorname{ann} x$, $\operatorname{ann} y$, and $\operatorname{ann} z$, then the vertices $[x]$, $[y]$, and $[z]$ in $\Gamma_E(R)$ are all adjacent to each other by Lemma 1.1, and therefore span a complete subgraph of $\Gamma_E(R)$. Hence $\Gamma_E(R)$ contains a 3-cycle, so $\mathrm{g}(\Gamma_E(R)) = 3$.

Case 2: If R has exactly one associated prime, $\operatorname{ann} y$, then every other vertex in $\Gamma_E(R)$ is adjacent to $[y]$ by Lemma 1.1. If there is any cycle in $\Gamma_E(R)$, then there are some vertices $[x_1]$, $[x_2]$ distinct from $[y]$ with $[x_1] - [x_2]$. But these are both adjacent to $[y]$, creating the 3-cycle $[y] - [x_1] - [x_2] - [y]$. So $\mathrm{g}(\Gamma_E(R)) = 3$.

Case 3: Now assume that R has exactly 2 associated primes, and let $\operatorname{ass} R = \{\operatorname{ann} x, \operatorname{ann} y\}$. Let $[x_1]$ and $[x_2]$ be two vertices distinct from $[x]$ and $[y]$ such that $[x_1] - [x_2]$. By Lemma 1.1, $[x_1]$ is adjacent to an associated prime. Without loss of generality, let $[x_1] - [x]$. Also, $[x_2]$ is adjacent to either $[x]$ or $[y]$. In the first case, we have a 3-cycle $[x] - [x_1] - [x_2] - [x]$ and in the second case, we have a 4-cycle $[x] - [x_1] - [x_2] - [y]$. Now assume that given any two vertices of $\Gamma_E(R)$, at least one is an associated prime. Let $[x_1] - [x_2] - \cdots - [x_n] - [x_1]$ be a cycle in $\Gamma_E(R)$ of minimal length, and let $n \geq 4$. Since at least one of $[x_1]$ and $[x_2]$ is an associated prime, without loss of generality let $[x_1]$ be an associated prime. Also, at least one of $[x_3]$ and $[x_4]$ is an associated prime. If $[x_3]$ is an associated prime, we have the 3-cycle $[x_1] - [x_2] - [x_3] - [x_1]$, and if $[x_4]$ is an associated prime, we have the 4-cycle $[x_1] - [x_2] - [x_3] - [x_4] - [x_1]$. □

The following corollary is a direct result of the proof of Theorem 3.1.

Corollary 3.2. *If $\Gamma_E(R)$ has girth 4, then R must have exactly 2 associated primes.*

The following proposition gives a relationship between the girths of the two graphs. Note that the inequality is opposite that of the diameter relationship stated in the previous section.

Proposition 3.3. *If $\Gamma_E(R)$ contains a cycle, then $\mathrm{g}(\Gamma_E(R)) \geq \mathrm{g}(\Gamma(R))$.*

Proof. Let $[x_1]-[x_2]-\cdots-[x_n]-[x_1]$ be a cycle in $\Gamma_E(R)$. For each $[x_i]$, choose one $y_i \in [x_i]$. Then by the definition of multiplication of our equivalence classes, $y_1-y_2-\cdots-y_n-y_1$ is a cycle in $\Gamma(R)$ of equal length. So $\mathrm{g}(\Gamma_E(R)) \geq \mathrm{g}(\Gamma(R))$. □

Corollary 3.4. *If* $\mathrm{g}(\Gamma_E)=3$, *then* $\mathrm{g}(\Gamma)=3$.

We now examine the situation in which $\Gamma_E(R)$ has girth 4 and conclude that it is impossible.

Theorem 3.5. *For R a commutative Noetherian ring*, $\mathrm{g}(\Gamma_E(R)) \neq 4$.

Proof.

Suppose that $\Gamma_E(R)$ has girth 4. By Corollary 3.2, R has exactly two associated primes, so let $\operatorname{ass} R=\{\operatorname{ann} x, \operatorname{ann} y\}$.

Since $\operatorname{ann} x$ and $\operatorname{ann} y$ are associated primes, $[x]-[y]$ by Lemma 1.1. Let $[z]$ be some other vertex of $\Gamma_E(R)$. Then $[z]$ must be adjacent to at least one of $[x]$ or $[y]$. But if it is adjacent to both $[x]$ and $[y]$ we have a 3-cycle, so $[z]$ is adjacent to exactly one of $[x]$ or $[y]$. Thus the vertex set of $\Gamma_E(R)$ minus $\{[x],[y]\}$ can be partitioned into two disjoint subsets, one adjacent to $[x]$ and one adjacent to $[y]$. We refer to these subsets as X and Y, respectively.

As mentioned earlier, since R is Noetherian, there is at least one maximal element of $\mathfrak{F}$, and this annihilator is an associated prime. Without loss of generality, let $\operatorname{ann} x$ be maximal in $\mathfrak{F}$. We claim that $\operatorname{ann} y$ is also maximal in $\mathfrak{F}$. Now if $\operatorname{ann} y \subseteq \operatorname{ann} w$ for some w, then $\operatorname{ann} w \subseteq \operatorname{ann} m$ for some maximal element $\operatorname{ann} m \in \mathfrak{F}$, but since $\operatorname{ann} m$ is an associated prime, $\operatorname{ann} m=\operatorname{ann} y$ or $\operatorname{ann} m=\operatorname{ann} x$. In the latter case, $\operatorname{ann} y \subseteq \operatorname{ann} x$, so $[x]$ and $[y]$ are both adjacent to a common vertex. This creates a 3-cycle, contradicting that $\mathrm{g}(\Gamma_E(R))=4$. So both $\operatorname{ann} y$ and $\operatorname{ann} x$ are maximal in $\mathfrak{F}$.

Suppose that $[x]^2=[0]$ and $[y]^2=[0]$, and consider the class $[x+y]$. This class is annihilated by both $[x]$ and $[y]$, so either $[x+y]=[0]$ or $[x+y]$ is in the vertex set of $\Gamma_E(R)$. If $[x+y]=[0]$, then $[x]=[y]$, contrary to our assumption. So $[x+y]$ is in the vertex set of $\Gamma_E(R)$. Since $[y]$ is adjacent to no vertex of X, $[x+y]\neq[x]$. Similarly, since $[x]$ is adjacent to no vertex of Y, $[x+y]\neq[y]$. So $\Gamma_E(R)$ contains the 3-cycle $[x+y]-[x]-[y]-[x+y]$, a contradiction.

Now suppose that $[x]^2\neq[0]$ and $[y]^2\neq[0]$. Then $\operatorname{ann} x \cap \operatorname{ann} y=\{0\}$. Now multiplying any $[x_j]\in X$ and $[y_i]\in Y$, we see that since $[x_j]\in \operatorname{ann} x$ and $[y_i]\in \operatorname{ann} y$, $[x_j y_i]\in \operatorname{ann} x \cap \operatorname{ann} y=\{0\}$. If we break up the vertex set of $\Gamma_E(R)$ into $X\cup\{[y]\}$ and $Y\cup\{[x]\}$, we see that $\Gamma_E(R)$ is complete bipartite, and $\Gamma_E(R)=K_{n,m}$ with $n, m\neq 1$, which contradicts Lemma 1.2.

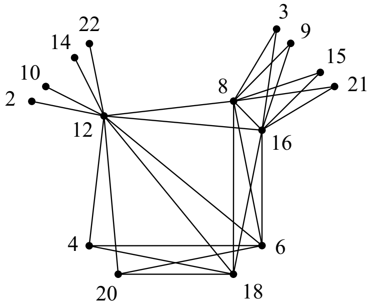
Without loss of generality, let $[x]^2=[0]$ and $[y]^2\neq[0]$. Let $[x]-[y]-[z]-[w]-[x]$ be a 4-cycle in $\Gamma_E(R)$, with $[w]\in X$, $[z]\in Y$. Then there is a 4-cycle $x-y-z-w-x$ in $\Gamma(R)$. By the previous discussion, $x^2=0$ and $y^2\neq 0$.

	diam $\gamma_E(R) =$		
diam $\gamma(R)$	∞	3	4
∞	$\mathbb{Z}_4$	impossible (Proposition 3.3)	
3	$\mathbb{Z}_{12}$	$\mathbb{Z}_{24}$	impossible (Theorem 3.5)
4	$\mathbb{Z}_{15}$	impossible (Corollary 3.2)	impossible (Theorem 3.5)

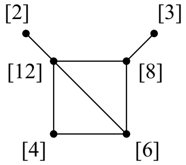
Table 2. Possibilities for $g(\Gamma(R))$ and $g(\Gamma_E(R))$, with examples.

Since $\operatorname{ann} y$ is maximal in the set of annihilators of R, there is some m in $\operatorname{ann} y$ but not in $\operatorname{ann} w$. Note that $mw \neq 0$, but $\operatorname{ann} mw \supseteq \{x, z, y\}$. Since $mw - y$ but $y^2 \neq 0$, $\operatorname{ann} mw \neq \operatorname{ann} y$. Also since mw is adjacent to both x and z, and x and z are not adjacent, $\operatorname{ann} mw \neq \operatorname{ann} x$ and $\operatorname{ann} mw \neq \operatorname{ann} z$. So we have the 3-cycles $x - y - mw - x$ and $z - y - mw - z$ that do not reduce under the equivalence relation. So $\Gamma_E(R)$ contains a 3-cycle and $g(\Gamma_E(R)) \neq 4$. □

We summarize with Table 2, which shows all possible combinations of girths for $\Gamma(R)$ and $\Gamma_E(R)$. We illustrate the case (3, 3) with the graphs of the ring $\mathbb{Z}_{24}$, which does not have shrinking girth:



$\Gamma(\mathbb{Z}_{24})$



$\Gamma_E(\mathbb{Z}_{24})$

4. Cut-vertices

In this section, we examine the properties of cut-vertices of $\Gamma_E(R)$. Since $\Gamma_E(R)$ is connected, the vertex $[a]$ is a cut-vertex of $\Gamma_E(R)$ exactly when removing the vertex $[a]$ and its incident edges causes $\Gamma_E(R)$ to no longer be connected.

We begin with an interesting result concerning cut-vertices and ideals of the ring. The following theorem is very similar to [Axtell et al. 2009, Theorem 4.4], which deals with cut-vertices of the original zero-divisor graph $\Gamma(R)$.

Theorem 4.1. *If $[a]$ is a cut-vertex of $\Gamma_E(R)$, then $[a] \cup \{0\}$ forms an ideal of R.*

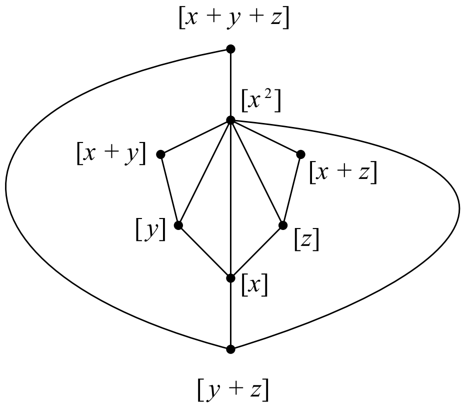
Proof. Let $[a]$ be a cut-vertex of $\Gamma_E(R)$ and let $[a]$ partition $\Gamma_E(R)$ into Γ_b and Γ_c. Let $[b] \in \Gamma_b$ with $[a] - [b]$ and $[c] \in \Gamma_c$ with $[a] - [c]$. Let $a_1, a_2 \in [a] \cup \{0\}$. Since $a_1 + a_2 \in \operatorname{ann} b \cap \operatorname{ann} c$, $a_1 + a_2 \in [a] \cup \{0\}$. If $r \in R$, then $c(ra) = r(ca) = 0$, so $ra \in \operatorname{ann} c$. Similarly, $ra \in \operatorname{ann} b$. So $ra \in \operatorname{ann} b \cap \operatorname{ann} c = [a] \cup \{0\}$. This shows that $[a] \cup \{0\}$ is an ideal of R. □

Theorem 4.2. *If $[a]$ is a cut-vertex of $\Gamma_E(R)$, then* $\operatorname{ann} a$ *is maximal in $\mathfrak{F}$.*

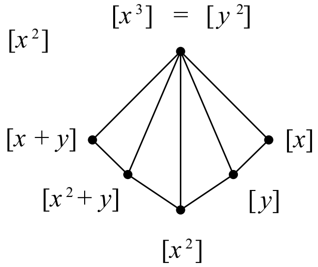
Proof. Let $[a]$ be a cut-vertex of $\Gamma_E(R)$, and let X and Y be mutually separated subgraphs of $\Gamma_E(R)$ with $V(X \cup Y) = V(\Gamma_E(R)) \setminus [a]$. Let $[x] \in X$ and $[y] \in Y$. Then for any $[x_1] \in X$ we have $y \in \operatorname{ann} a \smallsetminus \operatorname{ann} x_1$, and for any $[y_1] \in Y$ we have $x \in \operatorname{ann} a \smallsetminus \operatorname{ann} y_1$. Thus $\operatorname{ann} a \not\subseteq \operatorname{ann} x_1$ and $\operatorname{ann} a \not\subseteq \operatorname{ann} y_1$, and so $\operatorname{ann} a$ is maximal in $\mathfrak{F}$. □

The converse of this theorem does not hold. We may have $\operatorname{ann} x$ maximal in $\mathfrak{F}$, yet not have $[x]$ be a cut-vertex. For example, here are two equivalence graphs, one on 6 vertices and one on 8, each with no cut vertex:

$\Gamma_E(\mathbb{Z}_2[x, y, z] / (x^3, y^2, z^2, xy, xz))$



$\Gamma_E(\mathbb{Z}_2[x, y] / (x^4, xy, x^3 + y^2))$



Both of these rings contains an annihilator ideal which maximal in $\mathfrak{F}$, and therefore an associated prime.

The next corollary follows immediately from Theorem 4.2:

Corollary 4.3. *If $[a]$ is a cut-vertex of $\Gamma_E(R)$, then* $\operatorname{ann} a$ *is an associated prime.*

Theorem 4.4. *If $[a]$ is a cut-vertex of $\Gamma_E(R)$, then all other associated primes of $\Gamma_E(R)$ are contained in only one connected component of $\Gamma_E(R) \setminus [a]$.*

Proof. Suppose that X and Y are two mutually separated connected components of $\Gamma_E(R) \setminus [a]$, and that each contains an associated prime. By Lemma 1.1, these associated primes are adjacent, and so X and Y are connected, a contradiction. □

Theorem 4.5. *If $\Gamma_E(R)$ has at least 2 cut-vertices, then it has diameter 3.*

Proof. Let $[a]$ and $[b]$ be cut-vertices of $\Gamma_E(R)$. Since $[a]$ is a cut-vertex, there is some $[x_a]$ such that any path connecting $[x_a]$ and $[b]$ must include $[a]$. Similarly, since $[b]$ is a cut-vertex, there is some $[x_b]$ such that any path connecting $[x_b]$ and $[a]$ must include $[b]$. Therefore any path from $[x_a]$ to $[x_b]$ must include both $[a]$ and $[b]$ and so $d([a], [b]) \geq 3$. Since $\Gamma_E(R)$ is connected, $\operatorname{diam} \Gamma_E(R) = 3$. □

Acknowledgements

This paper was written during the Research Experience for Undergraduates conducted at Brigham Young University in the summer of 2010. This research was funded by the National Science Foundation (DMS-0453421) and Brigham Young University. We would like to especially acknowledge the help of fellow REU students Cathryn Holm and Kaylee Kooiman, as well as TAs Chelsea Johnson and Donald Sampson.

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

[Axtell et al. 2009] M. Axtell, J. Stickles, and W. Trampbachls, "Zero-divisor ideals and realizable zero-divisor graphs", *Involve* **2**:1 (2009), 17–27. MR 2010b:13011 Zbl 1169.13301

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

[Mulay 2002] S. B. Mulay, "Cycles and symmetries of zero-divisors", *Comm. Algebra* **30**:7 (2002), 3533–3558. MR 2003j:13007a Zbl 1087.13500

[Spiroff and Wickham 2011] S. Spiroff and C. Wickham, "A zero divisor graph determined by equivalence classes of zero divisors", *Comm. Algebra* **39**:7 (2011), 2338–2348. MR 2821714 Zbl 1225.13007

Received: 2011-01-19 Revised: 2011-08-02 Accepted: 2011-08-18

blakej2@hotmail.com
Department of Mathematics, Utah Valley University,
Orem, UT 84058, United States

martine@william.jewell.edu
Department of Physics and Mathematics,
William Jewell College, Liberty, MO 64068, United States

new4@tcnj.edu
Department of Mathematics and Statistics,
The College of New Jersey, Ewing, NJ 08628, United States

dane.skabelund@gmail.com
Department of Mathematics, Brigham Young University,
Provo, UT 84602, United States

Total positivity of a shuffle matrix

Audra McMillan

(Communicated by John C. Wierman)

Holte introduced a $n \times n$ matrix P as a transition matrix related to the carries obtained when summing n numbers base b. Since then Diaconis and Fulman have further studied this matrix proving it to also be a transition matrix related to the process of b-riffle shuffling n cards. They also conjectured that the matrix P is totally nonnegative. In this paper, the matrix P is written as a product of a totally nonnegative matrix and an upper triangular matrix. The positivity of the leading principal minors for general n and b is proven as well as the nonnegativity of minors composed from initial columns and arbitrary rows.

1. Introduction

Holte [1997] introduced an $n \times n$ matrix P, with entries

$$P(i, j) = \frac{1}{b^n} \sum_{r=0}^{j-\lfloor i/b \rfloor} (-1)^r \binom{n+1}{r} \binom{n-1-i+(j+1-r)b}{n}$$

where the $P(i, j)$ entry gives the probability that when adding n random numbers base b, the next carry will be j, given that the previous carry was i. This matrix was then further studied in [2009a; Diaconis and Fulman 2009b], where it is noted that this is also a transition matrix related to card shuffling, where the $P(i, j)$ entry records the probability that a b-riffle shuffle of a permutation with i descents will lead to a permutation with j descents. Note that the rows and columns of this matrix are indexed by $0, \dots, n-1$.

Holte proved a number of properties of the matrix P, including that P has eigenvalues given by the geometric sequence $1, b^{-1}, \dots, b^{-(n-1)}$, implying that the determinant is positive for positive b.

A matrix will be referred to as *totally nonnegative* if every minor is nonnegative and *totally positive* if every minor is positive. Note that in some texts, such as [Pinkus 2010] and [Karlin 1968] these terms are replaced by *totally positive* and

MSC2010: 15B48, 60C05.

Keywords: total positivity, shuffle, minors.

strictly totally positive respectively. Totally nonnegative matrices figure prominently in a wide range of mathematical disciples including, but not limited to, combinatorics, stochastic processes and probability theory. Many properties of totally nonnegative matrices are known including eigenvalue/eigenvector properties and factorisation of such matrices. A good reference for the theory and applications of total positivity is [Pinkus 2010] and some further results on stochastic totally nonnegative matrices are included in [Gasca and Micchelli 1996].

Diaconis and Fulman [2009a, Remark after Lemma 4.2] conjectured that the matrix P is totally nonnegative for all positive integers n and b. Their paper included a proof that for all n and b, P is totally nonnegative of order 2, that is all the 2×2 minors are nonnegative, and that when b is a power of 2, P is totally nonnegative. Unfortunately, their method of proof does not generalise to other b. The aim of this paper is to make progress on the general conjecture.

Recall the following result:

Theorem 1. *Let $A = (a_{ij})$ be an $n \times n$ nonsingular matrix whose rows and columns are indexed by $0, \dots, n-1$. Then A is totally nonnegative if and only if A satisfies*

(i) $A\begin{pmatrix} 0, \dots, k-1 \\ 0, \dots, k-1 \end{pmatrix} > 0$ *for* $k = 1, \dots, n$,

(ii) $A\begin{pmatrix} i_1, \dots, i_k \\ 0, \dots, k-1 \end{pmatrix}$ *for* $0 \le i_1 < \cdots < i_k \le n-1$ *and* $k = 1, \dots, n$,

(iii) $A\begin{pmatrix} 0, \dots, k-1 \\ j_1, \dots, j_k \end{pmatrix}$ *for* $0 \le j_1 < \cdots < j_k \le n-1$ *and* $k = 1, \dots, n$,

where $A\begin{pmatrix} i_1, \dots, i_k \\ j_1, \dots, j_k \end{pmatrix}$ *denotes the minor composed of rows* $i_1, \dots, i_k$ *and columns* $j_1, \dots, j_k$.

A proof of this can be found in [Pinkus 2010, Proposition 2.15].

In this paper, we will prove (i) and (ii) for the matrix P, hence reducing the conjecture to condition (iii). Proving these conditions hold for P is equivalent to proving that they hold for $P' = b^n P$, so this matrix will be dealt with instead.

2. Proof of total nonnegativity claims

Firstly, note that

$$P'(i, j) = \sum_{r=0}^{j-\lfloor i/b \rfloor} (-1)^r \binom{n+1}{r} \binom{n-1-i+(j+1-r)b}{n}$$

$$= \sum_{r=0}^{j} (-1)^r \binom{n+1}{r} \binom{n-1-i+(j+1-r)b}{n},$$

which implies

$$P' = \left(\binom{n-1-i+(j+1)b}{n}\right)_{\substack{0\le i\le n-1\\0\le j\le n-1}} \left((-1)^{j-i}\binom{n+1}{j-i}\right)_{\substack{0\le i\le n-1\\0\le j\le n-1}},$$

where $\binom{n}{k} = 0$ if $k < 0$. Let's call the first matrix A, and the second B. Note that B is upper unitriangular.

Using the Vandermonde convolution note that

$$\sum_{k=0}^{n-i} \binom{n-i}{n-i-k}\binom{jb}{i+k} = \binom{n-i+jb}{n},$$

so A can be further factored as

$$A = \left[\binom{n-i-1}{j-i}\right]_{i,j} \left[\binom{(j+1)b}{i+1}\right]_{i,j}.$$

Let's call these matrices C and D, respectively. Note that C is upper unitriangular, so this factorisation of P' implies that $\det P' = \det D$.

Lemma 2. *C is totally nonnegative.*

Proof. Obviously all the leading principal minors of C are 1, and all other minors composed of k initial columns and k arbitrary rows are 0 since C is upper unitriangular.

Now let $k \in \mathbb{Z}$, $1 \le k \le n$ and $0 \le j_1 < \cdots < j_k \le n-1$. Again using the Vandermonde convolution we observe that

$$\sum_{p=0}^{k-i-1} \binom{k-i-1}{p}\binom{n-k}{j_{l+1}-i-p} = \binom{n-i-1}{j_{l+1}-i},$$

so

$$C\binom{0,\dots,k-1}{j_1,\dots,j_k} = \left|\left[\binom{n-i-1}{j_{l+1}-i}\right]_{i,l}\right| = \left|\left[\binom{k-i-1}{l-i}\right]_{i,l}\right| \left|\left[\binom{n-k}{j_{l+1}-i}\right]_{i,l}\right|$$
$$= \left|\left[\binom{n-k}{j_{l+1}-i}\right]_{i,l}\right|. \qquad (*)$$

A sequence $(a_i)_{0\le i<\infty}$ is called a Pólya frequency sequence of infinite order if the corresponding infinite kernel matrix

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & \cdots \\ 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

is totally nonnegative. The matrix (*) is nonnegative since it is a submatrix of the infinite kernel matrix of the sequence

$$\left(\binom{n-k}{0},\binom{n-k}{1},\dots,\binom{n-k}{n-k}\right),$$

which is a Pólya frequency sequence of infinite order according to the classification of Pólya frequency sequences in [Karlin 1968, Theorem 5.3, Chapter 8].

Therefore, by Theorem 1, matrix C is totally nonnegative for all n. □

Lemma 3. *D is totally nonnegative.*

Proof. D is a submatrix of the upper triangular Pascal matrix

$$\left[\binom{j}{i}\right]_{i,j}$$

which is simply the reflection of C about the antidiagonal where the dimension is $nb+1$, and hence is totally nonnegative [Pinkus 2010, Propositions 1.2 and 1.3]. Therefore D is totally nonnegative. □

Corollary 4. *A is totally nonnegative.*

Proof. Since the product of totally nonnegative matrices is totally nonnegative, A is totally nonnegative. □

Proposition 5. *Conditions* (*i*) *and* (*ii*) *of Theorem* 1 *hold for matrix P' for general n and b.*

Proof. Let $k\in\mathbb{Z}$, $1\le k\le n$ and $0\le i_1<\dots<i_k\le n-1$. From the Cauchy–Binet formula and the fact that B is upper unitriangular,

$$P'\begin{pmatrix} i_1,\dots,i_k\\ 0,\dots,k-1\end{pmatrix}=A\begin{pmatrix} i_1,\dots,i_k\\ 0,\dots,k-1\end{pmatrix}\ge 0$$

and

$$\begin{aligned}P'\begin{pmatrix}0,\dots,k-1\\0,\dots,k-1\end{pmatrix}&=A\begin{pmatrix}0,\dots,k-1\\0,\dots,k-1\end{pmatrix}\\&=\sum_{0\le m_1<\dots<m_k\le n-1} C\begin{pmatrix}0,\dots,k-1\\m_1,\dots,m_k\end{pmatrix}D\begin{pmatrix}m_1,\dots,m_k\\0,\dots,k-1\end{pmatrix}\\&\ge D\begin{pmatrix}0,\dots,k-1\\0,\dots,k-1\end{pmatrix}=\left|\left[\binom{(j+1)b}{i+1}\right]_{i,j}\right|.\end{aligned}$$

Here the inequality follows from the fact that C and D are totally nonnegative and C is upper unitriangular.

However this is simply the determinant of a smaller version of D, with n replaced by k and therefore by the previous factorisation of P', this is equal to the determinant of the P' matrix of dimension k, which is positive (as stated earlier) so we are done. □

One might hope that condition (iii) could be proved similarly by noting that

$$P'\binom{0,\dots,k-1}{j_1,\dots,j_k} = \sum_{0\le m_1<\cdots<m_k\le n-1} A\binom{0,\dots,k-1}{m_1,\dots,m_k} B\binom{m_1,\dots,m_k}{j_1,\dots,j_k}.$$

However the proof of condition (ii) relied on the fact that the minors of B involved were clearly seen to be 0 or 1 so this equation easily simplified. This is not the case for the above equation since little has been established about general minors of B. Progress might still be made if all minors of size k were nonnegative for some k however small examples show this to be unlikely, for example this is not true for minors of size 2 for any n. If the conjecture is true, it seems likely that a new approach is required to prove condition (iii).

3. Acknowledgements

This work was supported by the University of Sydney School of Mathematics vacation scholarship. A special thanks is given to Dr. Anthony Henderson for his support with this project.

References

[Diaconis and Fulman 2009a] P. Diaconis and J. Fulman, "Carries, shuffling, and an amazing matrix", *Amer. Math. Monthly* **116**:9 (2009), 788–803. MR 2011d:60027 Zbl 1229.60011

[Diaconis and Fulman 2009b] P. Diaconis and J. Fulman, "Carries, shuffling, and symmetric functions", *Adv. in Appl. Math.* **43**:2 (2009), 176–196. MR 2010m:60028 Zbl 1172.60002

[Gasca and Micchelli 1996] M. Gasca and C. A. Micchelli (editors), *Total positivity and its applications* (Jaca, 1994), Mathematics and its Applications **359**, Kluwer, Dordrecht, 1996. MR 97f:00029 Zbl 0884.00045

[Holte 1997] J. M. Holte, "Carries, combinatorics, and an amazing matrix", *Amer. Math. Monthly* **104**:2 (1997), 138–149. MR 98g:15034 Zbl 0889.15021

[Karlin 1968] S. Karlin, *Total positivity, I*, Stanford University Press, Stanford, CA, 1968. MR 37 #5667 Zbl 0219.47030

[Pinkus 2010] A. Pinkus, *Totally positive matrices*, Cambridge Tracts in Mathematics **181**, Cambridge University Press, Cambridge, 2010. MR 2010k:15065 Zbl 1185.15028

Received: 2011-02-24 Revised: 2011-07-17 Accepted: 2011-09-04

amcm7623@uni.sydney.edu.au *School of Mathematics and Statistics, University of Sydney,*
Sydney, NSW 2006, Australia

Betti numbers of order-preserving graph homomorphisms

Lauren Guerra and Steven Klee

(Communicated by Jim Haglund)

For graphs G and H with totally ordered vertex sets, a function mapping the vertex set of G to the vertex set of H is an order-preserving homomorphism from G to H if it is nondecreasing on the vertex set of G and maps edges of G to edges of H. In this paper, we study order-preserving homomorphisms whose target graph H is the complete graph on n vertices. By studying a family of graphs called nonnesting arc diagrams, we are able to count the number of order-preserving homomorphisms (and more generally the number of order-preserving multihomomorphisms) mapping any fixed graph G to the complete graph K_n.

1. Introduction

The study of graph homomorphisms has been the subject of a great deal of recent work in the fields of enumerative, algebraic, and topological combinatorics. The recent survey [Borgs et al. 2006] is an excellent source on the many facets of enumerating graph homomorphisms, while [Kozlov 2008] outlines a more topological approach. In this paper, we study combinatorial properties of order-preserving homomorphisms between two graphs G and H as introduced by Braun, Browder and Klee [Braun et al. 2011].

Throughout this paper, $V(G)$ and $E(G)$ will denote the vertex set and edge set respectively of a graph G. All graphs are assumed to be simple, meaning that loops and multiple edges are not allowed.

Let G be a graph on vertex set $[m] = \{1, 2, \dots, m\}$ and let H be a graph on vertex set $\{x_1, x_2, \dots, x_n\}$. We order the vertex set of G naturally, and we order the vertex set of H by declaring that $x_1 < x_2 < \cdots < x_n$. An *order-preserving homomorphism* from G to H is a function $\varphi : V(G) \to V(H)$ such that

(1) if $1 \le i < j \le m$, then $\varphi(i) \le \varphi(j)$, and

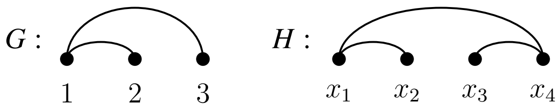
(2) if $(i, j) \in E(G)$, then $(\varphi(i), \varphi(j)) \in E(H)$.

MSC2010: primary 13D02; secondary 05A18, 06A06, 05C30.

Keywords: graph homomorphisms, Betti numbers, nonnesting partitions.

An order-preserving homomorphism $\varphi : G \to H$ may be presented as a vector $[\varphi(i)]_{i=1}^m = [\varphi(1), \dots, \varphi(m)]$.

Example 1.1. Let G and H be as follows:



Define functions $\varphi_1, \varphi_2, \varphi_3 : V(G) \to V(H)$ by

$$\varphi_1 : [x_1, x_2, x_2], \quad \varphi_2 : [x_1, x_2, x_4], \quad \varphi_3 : [x_1, x_3, x_4].$$

The functions φ_1 and φ_2 are order-preserving homomorphisms from G to H. Notice that since $(2, 3)$ is not an edge in G, having $\varphi_1(2) = \varphi_1(3)$ does not violate the definition of an order-preserving homomorphism. The function φ_3 is order-preserving, but it is not a homomorphism since $(1, 2) \in E(G)$, but $(\varphi(1), \varphi(2)) = (x_1, x_3) \notin E(H)$.

Rather than view each order-preserving homomorphism from G to H as a single function, it is often more convenient to encode several homomorphisms as a single object. An *(order-preserving) multihomomorphism* from G to H is a function $\eta : V(G) \to 2^{V(H)} \setminus \varnothing$ with the property that $[\varphi(i)]_{i=1}^m$ is an order-preserving homomorphism from G to H for all possible choices of $\varphi(i) \in \eta(i)$ and $1 \le i \le m$. The *complex of order-preserving homomorphisms* from G to H, denoted $\mathrm{OHOM}(G, H)$, is the collection of all multihomomorphisms from G to H.

For any graphs G and H, there is a geometric cell complex corresponding to $\mathrm{OHOM}(G, H)$ whose faces are labeled by multihomomorphisms from G to H. While the geometry of $\mathrm{OHOM}(G, H)$ is very interesting in its own right, it is not the primary focus of this paper, and we will not spend any further time discussing it. For reasons that are motivated by this underlying geometry, we define the *dimension* of a multihomomorphism $\eta \in \mathrm{OHOM}(G, H)$ to be

$$\dim \eta := \sum_{i=1}^{m} (|\eta(i)| - 1).$$

A zero-dimensional multihomomorphism is an order-preserving homomorphism. In this paper, we are primarily interested in a family of combinatorial invariants of $\mathrm{OHOM}(G, H)$ called its Betti numbers.

Definition 1.2. The *r-th Betti number* of the complex $\mathrm{OHOM}(G, H)$, denoted $\beta_r(G, H)$, counts the number of multihomomorphisms $\eta \in \mathrm{OHOM}(G, H)$ with $\dim \eta = r$.

Example 1.3. Let G and H be as in Example 1.1. The following table encodes a one-dimensional multihomomorphism $\eta \in \mathrm{OHOM}(G, H)$:

$\eta(1)$	$\eta(2)$	$\eta(3)$
x_1	x_2	x_2
		x_4

The two distinct choices of elements $[\varphi(1), \varphi(2), \varphi(3)]$ correspond to the order-preserving homomorphisms φ_1 and φ_2 of Example 1.1.

The following proposition is a consequence of from our definitions of order-preserving homomorphisms. We introduce the following notation, which will be used for the remainder of the paper. If X and Y are subsets of some totally ordered set (for our purposes, either $[m]$ or $\{x_1, \ldots, x_n\}$), we write $X \leq Y$ (or $X < Y$) to indicate that $x \leq y$ (similarly $x < y$) for all $x \in X$ and all $y \in Y$.

Proposition 1.4. *Let G and H be graphs with*

$$V(G) = [m] \quad \textit{and} \quad V(H) = \{x_1, \ldots, x_n\}.$$

If $\eta \in \mathrm{OHOM}(G, H)$, then $\eta(1) \leq \eta(2) \leq \cdots \leq \eta(m)$. Moreover, if (i, j) is an edge in G, then $\eta(i) < \eta(j)$.

The purpose of this paper is to determine the Betti numbers $\beta_r(G, K_n)$ of the complex of order-preserving homomorphisms between a fixed graph G and the complete graph on n vertices. In order to more easily compute the Betti numbers $\beta_r(G, K_n)$, we use the following series of reductions outlined in [Braun et al. 2011, Section 5]. All relevant definitions are deferred to Section 2.

(1) We show that for any graph G, there is a nonnesting partition $\mathcal{P}$ of $[m]$ and a corresponding graph $\Gamma_{\mathcal{P}}$ on $[m]$, called an *arc diagram*, such that

$$\mathrm{OHOM}(G, K_n) = \mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n).$$

(2) We define a weight function $\omega_r(\Gamma_{\mathcal{P}}, K_n)$ that counts the number of r-dimensional multihomomorphisms in $\mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n)$ "minimally" determined by $\mathcal{P}$. These weights are ultimately easier to compute than the Betti numbers of $\mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n)$.

(3) We define a partial order, denoted $\preceq$, on the family of nonnesting partitions of $[m]$ and show that

$$\beta_r(\Gamma_{\mathcal{P}}, K_n) = \sum_{\mathcal{Q} \preceq \mathcal{P}} \omega_r(\Gamma_{\mathcal{Q}}, K_n).$$

In Section 3, we provide an explicit (and simple) closed formula for the weight function $\omega_r(\Gamma_{\mathcal{P}}, K_n)$ for any nonnesting partition $\mathcal{P}$.

2. Nonnesting partition graphs

Nonnesting partitions. A *partition* $\mathcal{P} = \{P_1, \dots, P_t\}$ of the set $[m]$ is a collection of nonempty subsets $P_i \subseteq [m]$ (called *blocks*) such that $P_i \cap P_j = \varnothing$ for all $i \neq j$ and $P_1 \cup \cdots \cup P_t = [m]$. We say that two blocks P_i and P_j *nest* if there exist $1 \leq a < b < c < d \leq m$ with $\{a, d\} \subseteq P_i$ and $\{b, c\} \subseteq P_j$ and there does not exist $e \in P_i$ with $b < e < c$. If no pair of blocks of $\mathcal{P}$ nest, we say that $\mathcal{P}$ is a *nonnesting partition* of $[m]$. The family of nonnesting partitions was originally introduced and studied by Postnikov; see [Reiner 1997, Remark 2].

Example 2.1. The partition $\mathcal{P}_1 = \{\{1, 4\}, \{2, 5, 6\}, \{3\}\}$ of [6] is a nonnesting partition. The partition $\mathcal{P}_2 = \{\{1, 3, 5\}, \{2, 6\}, \{4\}\}$ is nesting since the blocks $\{1, 3, 5\}$ and $\{2, 6\}$ nest.

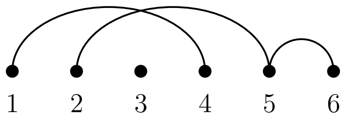
It is more illuminating to represent a partition $\mathcal{P}$ of $[m]$ as a graph $\Gamma_{\mathcal{P}}$ as follows.

Definition 2.2. Let $\mathcal{P}$ be a partition of $[m]$ and let $P_i = \{i_1, \dots, i_k\}$ be a block of $\mathcal{P}$ with $i_1 < \cdots < i_k$. The *arc diagram* $\Gamma_{\mathcal{P}}$ is the graph on vertex set $[m]$ whose edges are given by (i_j, i_{j+1}) for consecutive elements of P_i taken over all blocks of $\mathcal{P}$.

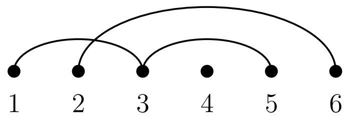
The name "arc diagram" is natural when the graph $\Gamma_{\mathcal{P}}$ is drawn so that its vertices are placed in a line and its edges are drawn as upper semicircular arcs, as shown in Example 2.3. In this representation, a partition $\mathcal{P}$ is nonnesting exactly when no arc of $\Gamma_{\mathcal{P}}$ is nested below another.

Example 2.3. Let $\mathcal{P}_1 = \{\{1, 4\}, \{2, 5, 6\}, \{3\}\}$ and $\mathcal{P}_2 = \{\{1, 3, 5\}, \{2, 6\}, \{4\}\}$ be the partitions of [6] discussed in Example 2.1. The arc diagrams $\Gamma_{\mathcal{P}_1}$ and $\Gamma_{\mathcal{P}_2}$ are as follows:

$\{\{1, 4\}, \{2, 5, 6\}, \{3\}\}$



$\{\{1, 3, 5\}, \{2, 6\}, \{4\}\}$



The next proposition shows that in order to compute Betti numbers $\beta_r(G, K_n)$ for arbitrary graphs G, we need only study the Betti numbers of nonnesting arc diagrams.

Proposition 2.4 [Braun et al. 2011, Proposition 5.6]. *For any graph G on vertex set $[m]$, there exists a unique nonnesting partition $\mathcal{P}$ of $[m]$ such that $\Gamma_{\mathcal{P}}$ is a subgraph of G and* $\mathrm{OHOM}(G, K_n) = \mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n)$. *We call $\Gamma_{\mathcal{P}}$ the reduced arc diagram for G.*

Suppose there exist vertices $1 \leq a \leq b < c \leq d \leq m$ in G such that (a, d) and (b, c) lie in $E(G)$ (so that the edge (b, c) is nested below the edge (a, d)), and let G' be the graph obtained from G by removing the edge (a, d). The proof of Proposition 2.4 uses the observation that $\mathrm{OHOM}(G, K_n) = \mathrm{OHOM}(G', K_n)$ so that the reduced graph $\Gamma_{\mathcal{P}}$ is obtained from G by inductively removing the "top" arc in any pair of nested edges in G.

The goal for the remainder of this section is to describe a natural partial order on the family of nonnesting partitions of $[m]$. We then describe how to use this partial order to compute the Betti numbers $\beta_r(\Gamma_{\mathcal{P}}, K_n)$ of an arc diagram. For further information on posets and definitions of any undefined terms, we refer the reader to [Stanley 1997].

Definition 2.5. The *m-th diagram poset*, denoted $\mathcal{D}_m = (\mathcal{D}_m, \preceq)$, is the poset whose elements are arc diagrams of nonnesting partitions of $[m]$, partially ordered by $\mathcal{P} \preceq \mathcal{Q}$ if every arc of $\mathcal{Q}$ lies above an arc of $\mathcal{P}$.

The minimal element of $\mathcal{D}_m$ is the path of length $m-1$ on $[m]$, and the maximal element of $\mathcal{D}_m$ is the empty graph.

For example, there are five nonnesting partitions of [3]:

$$\begin{aligned}\mathcal{P}_1 &= \{\{1\}, \{2\}, \{3\}\},\\ \mathcal{P}_2 &= \{\{1, 3\}, \{2\}\},\\ \mathcal{P}_3 &= \{\{1, 2\}, \{3\}\},\\ \mathcal{P}_4 &= \{\{1\}, \{2, 3\}\},\\ \mathcal{P}_5 &= \{\{1, 2, 3\}\}.\end{aligned}$$

Let $\Gamma_1, \dots, \Gamma_5$ denote their corresponding arc diagrams, as shown in Figure 1.

If $(P, \leq)$ is a poset, a subset $U \subseteq P$ is a *upper order ideal* if $y \in U$ whenever $x \in U$ and $y \geq x$. An upper order ideal $U \subseteq P$ is *principal* if there is an element

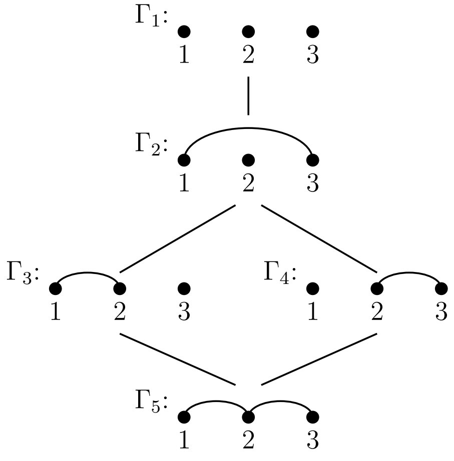


Figure 1. The Hasse diagram for $\mathcal{D}_3$.

$\alpha \in P$ such that $U = \{y \in P : y \geq \alpha\}$. The importance of the partial order on $\mathcal{D}_m$ is illustrated in the following proposition.

Proposition 2.6 [Braun et al. 2011, Proposition 5.8]. *If $\mathcal{P} \preceq \mathcal{Q}$ in $\mathcal{D}_m$, then*

$$\mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n) \subseteq \mathrm{OHOM}(\Gamma_{\mathcal{Q}}, K_n).$$

Further, for each multihomomorphism $\eta \in \mathrm{OHOM}(G_e, K_n)$, where G_e denotes the empty graph on vertex set $[m]$, the upper order ideal $U(\eta) \subseteq \mathcal{D}_m$ of arc diagrams whose OHOM *complexes contain η is principal.*

Proof. Fix a multihomomorphism $\eta \in \mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_n)$. We need to show that each choice $[\varphi(i) \in \eta(i)]_{i=1}^{m}$ yields an order-preserving homomorphism from $\Gamma_{\mathcal{Q}}$ to K_n so that $\eta \in \mathrm{OHOM}(\Gamma_{\mathcal{Q}}, K_n)$ as well.

Let (a, d) be an edge in $\Gamma_{\mathcal{Q}}$ with $a < d$. Since $\mathcal{P} \preceq \mathcal{Q}$, there is an edge (b, c) in $\Gamma_{\mathcal{P}}$ such that $a \leq b < c \leq d$. Since φ is an order-preserving homomorphism from $\Gamma_{\mathcal{P}}$ to K_n and (b, c) is an arc in $\Gamma_{\mathcal{P}}$, we see that $\varphi(a) \leq \varphi(b) < \varphi(c) \leq \varphi(d)$. The arc (a, d) was arbitrary, and hence $\varphi(a) < \varphi(d)$ for all arcs (a, d) in $\Gamma_{\mathcal{Q}}$. Thus φ is an order-preserving homomorphism from $\Gamma_{\mathcal{Q}}$ to K_n and $\eta \in \mathrm{OHOM}(\Gamma_{\mathcal{Q}}, K_n)$, as desired.

Suppose next that $\eta \in \mathrm{OHOM}(G_e, K_n)$. Consider the graph G on $[m]$ obtained as the union of all arc diagrams $\Gamma_{\mathcal{Q}}$ such that $\eta \in \mathrm{OHOM}(\Gamma_{\mathcal{Q}}, K_n)$, and let $\Gamma_{\mathcal{P}}$ denote the reduced arc diagram of G. Clearly $\mathcal{P} \preceq \mathcal{Q}$ for all nonnesting partitions $\mathcal{Q}$ whose OHOM complexes contain η. Thus $U(\eta)$ is generated by $\mathcal{P}$. □

Example 2.7. We illustrate Proposition 2.6 for the following multihomomorphism $\eta \in \mathrm{OHOM}(\Gamma_1, K_9)$, using the notation from Figure 1:

$\eta(1)$	$\eta(2)$	$\eta(3)$
x_1	x_4	x_7
x_3	x_6	x_9
	x_7	

Since $\eta(2) \cap \eta(3)$ is nonempty, the nonnesting partitions $\mathcal{P}$ for which η lies in $\mathrm{OHOM}(\Gamma_{\mathcal{P}}, K_9)$ are $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$. The corresponding graphs Γ_1, Γ_2, and Γ_3 form an upper order ideal in $\mathcal{D}_3$ that is generated by Γ_3.

Weights of nonnesting partition graphs. Proposition 2.6 gives a well defined notion of the minimal arc diagram $\Gamma_{\mathcal{Q}}$ whose OHOM complex supports a given multihomomorphism $\eta \in \mathrm{OHOM}(G_e, K_n)$. We make this more precise in the following definition.

Definition 2.8. Let $\mathcal{P}$ be a nonnesting partition of $[m]$. The *r-th weight* of $\mathcal{P}$ for n, denoted $\omega_r(\mathcal{P}, n)$, counts the number of r-dimensional multihomomorphisms $\eta \in \mathrm{OHOM}(G_e, K_n)$ such that $\mathcal{P}$ generates $U(\eta)$.

To be more specific, Proposition 2.6 says that for each nonnesting partition $\mathscr{Q}$ and each multihomomorphism $\eta \in \mathrm{OHOM}(\Gamma_{\mathscr{Q}}, K_n)$, there is a unique minimal nonnesting partition $\mathscr{P} \preceq \mathscr{Q}$ such that $\eta \in \mathrm{OHOM}(\Gamma_{\mathscr{P}}, K_n)$. This allows us to partition the r-dimensional multihomomorphisms of $\mathrm{OHOM}(\Gamma_{\mathscr{Q}}, K_n)$ according to the poset $\mathscr{D}_m$, as the following proposition indicates.

Proposition 2.9 [Braun et al. 2011, Proposition 5.10]. *For any nonnesting partition $\mathscr{Q}$,*

$$\beta_r(\Gamma_{\mathscr{Q}}, K_n) = \sum_{\mathscr{P} \preceq \mathscr{Q}} \omega_r(\mathscr{P}, n). \tag{2-1}$$

Recall that a collection of vertices W in a graph G is *independent* if there are no edges in G among the vertices in W. The following lemma provides a converse to Proposition 1.4 when computing weights.

Lemma 2.10. *Let η be a multihomomorphism of* $\mathrm{OHOM}(G_e, K_n)$, *and let $\mathscr{P}$ be the nonnesting partition whose arc diagram generates $U(\eta)$. Suppose $I = [a, c] \subseteq [m]$ is independent in $\Gamma_{\mathscr{P}}$. Then*

(1) $\eta(a) \cap \eta(c) \neq \varnothing$,

(2) $|\eta(a) \cap \eta(c)| = 1$, *and*

(3) *if $\eta(a) \cap \eta(c) = \{x_i\}$, then $\eta(b) = \{x_i\}$ for all $a < b < c$.*

Proof. To prove (1), suppose by way of contradiction that $\eta(a) \cap \eta(c) = \varnothing$. Consider the arc diagram $\Gamma_{\mathscr{Q}}$ obtained from $\Gamma_{\mathscr{P}}$ by adding the arc (a, c). Since I is independent in $\Gamma_{\mathscr{P}}$, the graph $\Gamma_{\mathscr{Q}}$ is the arc diagram of a nonnesting partition $\mathscr{Q}$.

First, we observe that $\mathscr{Q} \prec \mathscr{P}$ since $\Gamma_{\mathscr{P}}$ is a subgraph of $\Gamma_{\mathscr{Q}}$, and hence every arc of $\Gamma_{\mathscr{P}}$ lies above an arc of $\Gamma_{\mathscr{Q}}$. Next, we claim that $\eta \in \mathrm{OHOM}(\Gamma_{\mathscr{Q}}, K_n)$. Since (a, c) is the only edge in $E(\Gamma_{\mathscr{Q}}) \setminus E(\Gamma_{\mathscr{P}})$, we only need to check that (x, y) is an edge of K_n for any choice of $x \in \eta(a)$ and $y \in \eta(c)$. This follows immediately from our assumption that $\eta(a) \cap \eta(c) = \varnothing$.

Thus $\eta \in \mathrm{OHOM}(\Gamma_{\mathscr{Q}}, K_n)$ and $\mathscr{Q} \prec \mathscr{P}$, contradicting our assumption that the nonnesting partition $\mathscr{P}$ generates $U(\eta)$. This proves that $\eta(a) \cap \eta(c) \neq \varnothing$. Parts (2) and (3) follow immediately from the requirement that $\eta(a) \leq \eta(b) \leq \eta(c)$ for all $a < b < c$, together with the fact that $\eta(a) \cap \eta(c) \neq \varnothing$. □

Lemma 2.11 [Braun et al. 2011, Theorem 5.11]. *If $\Gamma_{\mathscr{P}}$ contains an arc (a, c) where $c - a > 2$, then $\omega_r(\mathscr{P}, n) = 0$.*

Proof. Suppose to the contrary that $\Gamma_{\mathscr{P}}$ contains such an arc and that $\omega_r(\mathscr{P}, n) \neq 0$. Let η be an r-dimensional multihomomorphism of $\mathrm{OHOM}(G_e, K_n)$ such that $\Gamma_{\mathscr{P}}$ generates $U(\eta)$.

Consider the intervals $I = [a, c-1]$ and $I' = [a+1, c]$. Since $\mathscr{P}$ is nonnesting,

I and I' are independent in $\Gamma_{\mathcal{P}}$. By Lemma 2.10, there is an element

$$x_i \in \eta(a) \cap \eta(c-1)$$

and moreover, $\eta(b) = \{x_i\}$ for all $a < b < c-1$. In particular, $\eta(a+1) = \{x_i\}$ since $a+1 < c-1$. By applying Lemma 2.10 to the interval I', we see that $\eta(a+1) \cap \eta(c) \neq \varnothing$ and hence $x_i \in \eta(c)$. Thus $x_i \in \eta(a) \cap \eta(c)$, which contradicts Proposition 1.4. □

Following [Braun et al. 2011], we call an arc diagram $\Gamma_{\mathcal{P}}$ containing no arcs of the form (i, j) with $j - i > 2$ a *small arc diagram*, and we say that the corresponding nonnesting partition $\mathcal{P}$ is a *small nonnesting partition.* In light of Lemma 2.11, we need only compute the weights $\omega_r(\mathcal{P}, K_n)$ for which $\Gamma_{\mathcal{P}}$ is a small arc diagram. The following two results are interesting enumerative results in their own right.

Proposition 2.12 ([Stanley 1997]). *The number of nonnesting arc diagrams on* $[m]$ *is enumerated by the* m*-th Catalan number*

$$C_m = \frac{1}{m+1}\binom{2m}{m}.$$

Proposition 2.13 ([Braun et al. 2011, Theorem 5.12]). *Let* F_m *be the* m*-th Fibonacci number with* $F_0 = F_1 = 1$. *The number of small arc diagrams on* $[m]$ *is* F_{2m-2}.

An example. As a more complicated example, we exhibit the weights and corresponding Betti numbers for all nonnesting partitions of $\{1, 2, 3\}$. We recall the arc diagrams $\Gamma_1, \ldots, \Gamma_5$ used in Figure 1.

Proposition 2.14. *For all* $r, n \geq 0$,

$$\omega_r(\Gamma_1, K_n) = \binom{n}{r+1}(r+1).$$

Proof. Let $\eta \in \mathrm{OHOM}(\Gamma_1, K_n)$ be a multihomomorphism whose upper order ideal $U(\eta)$ is generated by Γ_1. By Lemma 2.10, there is a single element $x_i \in \eta(1) \cap \eta(3)$ and $\eta(2) = \{x_i\}$. In order to compute $\omega_r(\Gamma_1, K_n)$, we first determine that there are $r+1$ distinct elements in $\eta(1) \cup \eta(2) \cup \eta(3)$. Indeed, by the inclusion-exclusion principle,

$$\begin{aligned}
&\left|\eta(1) \cup \eta(2) \cup \eta(3)\right| \\
&\quad = \left|\eta(1)\right| + \left|\eta(2)\right| + \left|\eta(3)\right| - \left|\eta(1) \cap \eta(2)\right| - \left|\eta(1) \cap \eta(3)\right| - \left|\eta(2) \cap \eta(3)\right| \\
&\qquad\qquad + \left|\eta(1) \cap \eta(2) \cap \eta(3)\right| \\
&\quad = (r+3) - 3 + 1 = r+1.
\end{aligned}$$

In order to describe any such multihomomorphism η, we must choose a subset $X \subseteq \{x_1, \dots, x_n\}$ of the $r+1$ distinct elements in $\eta(1) \cup \eta(2) \cup \eta(3)$, together with the single element $x_i \in X$ that is common to all three sets. Certainly there are $\binom{n}{r+1}(r+1)$ ways to make these choices. Having chosen X and $x_i \in X$, we take

$$\eta(1) = \{x \in X : x \leq x_i\}, \quad \eta(2) = \{x_i\}, \quad \text{and} \quad \eta(3) = \{x \in X : x \geq x_i\}. \quad \square$$

Proposition 2.15. *For all $r, n \geq 0$,*

$$\omega_r(\Gamma_2, K_n) = \binom{n}{r+1}\binom{r+1}{2}.$$

Proof. Let $\eta \in \mathrm{OHOM}(\Gamma_2, K_n)$ be an r-dimensional multihomomorphism whose upper order ideal $U(\eta)$ is generated by Γ_2. By Lemma 2.10, there is an element $x_i \in \eta(1) \cap \eta(2)$ and another element $x_j \in \eta(2) \cap \eta(3)$. Moreover, by Proposition 1.4, $\eta(1) \cap \eta(3) = \varnothing$ and hence $x_i \neq x_j$. Thus by the inclusion-exclusion principle, there are $r+1$ distinct elements in $\eta(1) \cup \eta(2) \cup \eta(3)$.

In order to describe any such multihomomorphism η, we must first choose a subset $X \subseteq \{x_1, \dots, x_n\}$ of the $r+1$ elements in $\eta(1) \cup \eta(2) \cup \eta(3)$, together with the elements $x_i \in \eta(1) \cap \eta(2)$ and $x_j \in \eta(2) \cap \eta(3)$. Certainly there are $\binom{n}{r+1}\binom{r+1}{2}$ ways to make these choices. Given the set X and distinguished elements x_i and x_j, we take

$$\eta(1) = \{x \in X : x \leq x_i\}, \quad \eta(2) = \{x \in X : x_i \leq x \leq x_j\}, \quad \eta(3) = \{x \in X : x \geq x_j\}. \quad \square$$

Proposition 2.16. *For all $r, n \geq 0$,*

$$\omega_r(\Gamma_3, K_n) = \binom{n}{r+2}\binom{r+2}{2}.$$

Proof. Let $\eta \in \mathrm{OHOM}(\Gamma_3, K_n)$ be an r-dimensional multihomomorphism whose upper order ideal $U(\eta)$ is generated by Γ_3. By Lemma 2.10, there is an element $x_j \in \eta(2) \cap \eta(3)$, and by Proposition 1.4, $\eta(1) \cap \eta(2) = \varnothing$. By the inclusion-exclusion principle, there are $r+2$ distinct elements in $\eta(1) \cup \eta(2) \cup \eta(3)$.

In order to describe any such multihomomorphism η, we must first choose a subset $X \subseteq \{x_1, \dots, x_n\}$ of the $r+2$ distinct elements in $\eta(1) \cup \eta(2) \cup \eta(3)$, together with the element $x_j \in \eta(2) \cap \eta(3)$ and the largest element x_i in $\eta(1)$. Certainly there are $\binom{n}{r+2}\binom{r+2}{2}$ ways to make these choices. As before, having chosen X, x_i and x_j, we take

$$\eta(1) = \{x \in X : x \leq x_i\}, \quad \eta(2) = \{x \in X : x_i < x \leq x_j\}, \quad \eta(3) = \{x \in X : x \geq x_j\}. \quad \square$$

Proposition 2.17. *For all $r, n \geq 0$,*

$$\omega_r(\Gamma_4, K_n) = \binom{n}{r+2}\binom{r+2}{2}.$$

Proof. The proof of this proposition follows by an argument that is symmetric to the one given to compute the weights $\omega_r(\Gamma_3, K_n)$. □

Proposition 2.18. *For all $r, n \geq 0$,*

$$\omega_r(\Gamma_5, K_n) = \binom{n}{r+3}\binom{r+2}{2}.$$

Proof. Let $\eta \in \mathrm{OHOM}(\Gamma_5, K_n)$ be an r-dimensional multihomomorphism whose upper order ideal $U(\eta)$ is generated by Γ_5. By Proposition 1.4, $\eta(1) \cap \eta(2)$, $\eta(2) \cap \eta(3)$, and $\eta(1) \cap \eta(3)$ are empty. Thus by the inclusion-exclusion principle, $|\eta(1) \cup \eta(2) \cup \eta(3)| = r+3$.

In order to describe such a multihomomorphism η, we must choose a subset $X \subseteq \{x_1, \ldots, x_n\}$ of the $r+3$ distinct elements of $\eta(1) \cup \eta(2) \cup \eta(3)$ together with the maximal elements x_i and x_j of $\eta(1)$ and $\eta(2)$ respectively. Having made these choices, we take

$$\eta(1) = \{x \in X : x \leq x_i\}, \quad \eta(2) = \{x \in X : x_i < x \leq x_j\}, \quad \eta(3) = \{x \in X : x > x_j\}.$$

Since $\eta(3)$ must be nonempty, we cannot choose x_j to be the maximal element of X. The number of ways to choose X, x_i, and x_j is $\binom{n}{r+3}\binom{r+2}{2}$, which completes the proof. □

3. Enumerative results

Our goal for this section is to prove the promised formula computing the weights $\omega_r(\mathcal{P}, n)$ for any small nonnesting partition $\mathcal{P}$. Before stating the main theorem, we establish notation that will be used for the remainder of the paper.

Proposition 3.1. *For any small nonnesting partition $\mathcal{P}$ of $[m]$, there is a unique constant $k = k(\mathcal{P})$ and a unique decomposition of $[m]$ into intervals $I_1, \ldots, I_k$ satisfying the following conditions.*

(P1) $I_1 \cup \cdots \cup I_k = [m]$,

(P2) $I_1 \leq I_2 \leq \cdots \leq I_k$,

(P3) $|I_j| \geq 2$ *for all j, and*

(P4) *each interval I_j satisfies exactly one of the following conditions*:

(i) *I_j is a maximal interval (under inclusion) that is independent in $\Gamma_{\mathcal{P}}$.*

(ii) *$I_j = \{i_j, i_{j+1}\}$ and (i_j, i_{j+1}) is an edge of $\Gamma_{\mathcal{P}}$.*

Proof. We induct on m. The result is clear when $m = 2$. When $m \geq 3$, we examine two cases.

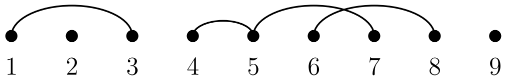
If $(1, 2)$ is an arc in $\Gamma_{\mathcal{P}}$, let $I_1 = \{1, 2\}$. Inductively, we may decompose the restriction of $\mathcal{P}$ to $[2, m]$ into intervals $I_2, \ldots, I_k$ satisfying conditions (P1)–(P4).

On the other hand, if $(1, 2)$ is not an arc in $\Gamma_{\mathcal{P}}$, let t be the largest element of $[m]$ such that $[1, t]$ is independent in $\Gamma_{\mathcal{P}}$. Let $I_1 = [1, t]$; if $t = m$, we have found the desired decomposition. Otherwise, if $t < m$, the restriction of $\Gamma_{\mathcal{P}}$ to $[t, m]$ is a small arc diagram, and we may inductively decompose the restriction of $\Gamma_{\mathcal{P}}$ to $[t, m]$ into intervals $I_2, \dots, I_k$ satisfying conditions (P1)–(P4).

In either of the above cases, we must check that the resulting interval decomposition $[m] = I_1 \cup \cdots \cup I_k$ satisfies conditions (P1)–(P4). Conditions (P1)–(P3) are satisfied by the inductive hypothesis. We must check, however, that if I_1 and I_2 are both edgefree as in condition (P4.i), then both are maximal under inclusion. By our construction, $I_1 = [1, t]$ is maximal. Since $t + 1 \notin I_1$ and $\mathcal{P}$ is small, either $(t, t + 1)$ or $(t - 1, t + 1)$ is an edge in $\Gamma_{\mathcal{P}}$. If $(t, t + 1)$ is an edge in $\Gamma_{\mathcal{P}}$, then $I_2 = \{t, t + 1\}$ satisfies condition (P4.ii). If $(t - 1, t + 1)$ is an edge in $\Gamma_{\mathcal{P}}$, then I_2 satisfies condition (P4.i), and $t - 1$ cannot be added to I_2 without violating the independence condition. Thus I_2 is maximal under inclusion, which completes the proof. □

Example 3.2. Consider the small arc diagram $\Gamma_{\mathcal{P}}$ for

$$\mathcal{P} = \{\{1, 3\}, \{4, 5, 7\}, \{6, 8\}, \{9\}\} :$$



The interval decomposition of $\Gamma_{\mathcal{P}}$ is

$$\begin{aligned} I_1 &= \{1, 2\}, \quad I_2 = \{2, 3, 4\}, \quad I_3 = \{4, 5\}, \\ I_4 &= \{5, 6\}, \quad I_5 = \{6, 7\}, \quad I_6 = \{7, 8, 9\}. \end{aligned}$$

Theorem 3.3. *Let $\mathcal{P}$ be a small nonnesting partition of $[m]$ with interval decomposition $I_1, \dots, I_k$ as described by Proposition 3.1. For any $r, n \geq 0$,*

$$\omega_r(\mathcal{P}, n) = \begin{cases} \binom{n}{l}\binom{l-1}{k} & \text{if } (1, 2), (m - 1, m) \in E(\Gamma_{\mathcal{P}}); \\ \binom{n}{l}\binom{l}{k} & \text{otherwise,} \end{cases} \tag{3-1}$$

where $l := r + m - \sum_{j \in J}(|I_j| - 1)$ and $J \subseteq [k]$ indexes those intervals described by condition (P4.i).

Proof. Fix a small nonnesting partition $\mathcal{P}$ of $[m]$. For each $1 \leq j \leq k$, let $I_j = [a_j, c_j]$. For any r-dimensional multihomomorphism $\eta \in \mathrm{OHOM}(\Gamma_e, K_n)$, we observe that $\sum_{i=1}^{m} |\eta(i)| = r + m$. If the arc diagram for $\Gamma_{\mathcal{P}}$ generates $U(\eta)$, then Lemma 2.10 prescribes the combinatorial structure of the intersections of the sets $\eta(i)$ within each interval $I_1, \cdots, I_k$.

As a consequence of these lemmas, we claim that as a *set*,

$$l := |\eta(1) \cup \cdots \cup \eta(m)| = r + m - \sum_{j \in J} (|I_j| - 1),$$

where $J \subseteq [k]$ indexes those intervals described by condition (P4.i). To see this, we simply observe that for each interval I_j with $j \in J$, there is a single element x_j common to the sets among $\{\eta(p) : p \in I_j\}$. When computing $|\eta(1) \cup \cdots \cup \eta(m)|$, each of these elements x_j is overcounted $|I_j| - 1$ times.

Thus in order to describe such a multihomomorphism η, we must first choose a subset $X \subseteq \{x_1, \dots, x_n\}$ of the l distinct elements of $\eta(1) \cup \cdots \cup \eta(m)$. This can be accomplished in $\binom{n}{l}$ ways.

Now suppose that $(1, 2)$ is not an arc of $\Gamma_{\mathcal{P}}$. The binomial coefficient $\binom{l}{k}$ counts the number of ways in which we may decompose the set X into pairwise disjoint intervals $A_0 < A_1 < \cdots < A_k$ so that the sets $A_1, \dots, A_k$ are nonempty. This follows from a standard stars-and-bars argument [Stanley 1997, Section 1.2] by arranging the elements of X linearly as

$$x_{i_1} \quad x_{i_2} \quad \cdots \quad x_{i_{l-1}} \quad x_{i_l},$$

with $i_1 < \cdots < i_l$ and choosing k of the spaces between consecutive elements of X to partition the set. This includes the possibility of choosing the space to the left of x_{i_1}, which corresponds to the case that A_0 is empty.

We now exhibit a bijection between the family of stars-and-bars partitions of X described in the previous paragraph and the collection of multihomomorphisms $\eta \in \mathrm{OHOM}(G_e, K_n)$ such that $\eta(1) \cup \cdots \cup \eta(m) = X$ and $\mathcal{P}$ generates $U(\eta)$.

Given pairwise disjoint intervals $A_0 < A_1 < \cdots < A_k$ that partition X with $A_1, \dots, A_k$ nonempty, let m_i denote the smallest element of A_i for $1 \leq i \leq k$. We determine the sets $\eta(i)$ by declaring that

- $A_0 \subseteq \eta(1)$,
- $A_j \subseteq \eta(c_j)$ for all $1 \leq j \leq k$, and
- $m_j \in \eta(b)$ for all $b \in [a_j, c_j]$ and all $j \in J$.

Lemma 2.10 and Proposition 1.4 show that this is a bijective correspondence. By symmetry, the same argument applies to the situation that $(m-1, m) \notin \Gamma_{\mathcal{P}}$.

In the case that both $(1, 2)$ and $(m-1, m)$ are edges in $\Gamma_{\mathcal{P}}$, an analogous bijection holds, with the exception that $\binom{l-1}{k}$ counts the number of partitions of X into nonempty, pairwise disjoint intervals $B_0 < \cdots < B_k$. Here we must require that B_0 and B_k are nonempty, since they describe the elements of $\eta(1)$ and $\eta(m)$, respectively. □

Example 3.4. We illustrate the proof of Theorem 3.3. Let $\mathcal{P}$ be the small partition from Example 3.2. Suppose $l = 11$ and (for simplicity) that

$$\eta(1) \cup \cdots \cup \eta(9) = \{x_1, \ldots, x_{11}\}.$$

The stars-and-bars decomposition

$$x_1 \quad x_2 \quad | \quad x_3 \quad | \quad x_4 \quad x_5 \quad x_6 \quad | \quad x_7 \quad | \quad x_8 \quad | \quad x_9 \quad | \quad x_{10} \quad x_{11}$$

gives

$$\begin{aligned} &A_0 = \{x_1, x_2\}, \quad A_1 = \{x_3\}, \quad A_2 = \{x_4, x_5, x_6\}, \quad A_3 = \{x_7\}, \\ &A_4 = \{x_8\}, \qquad\quad A_5 = \{x_9\}, \quad A_6 = \{x_{10}, x_{11}\}. \end{aligned}$$

This, in turn corresponds to the following multihomomorphism η:

$\eta(1)$	$\eta(2)$	$\eta(3)$	$\eta(4)$	$\eta(5)$	$\eta(6)$	$\eta(7)$	$\eta(8)$	$\eta(9)$
x_1								
x_2								
x_3	x_3							
	x_4	x_4	x_4					
			x_5					
			x_6					
				x_7				
				x_8	x_8			
					x_9	x_9		
						x_{10}	x_{10}	x_{10}
								x_{11}

We have shaded the blocks $A_j \subseteq \eta(c_j)$ for all $1 \le j \le 6$, where the intervals $I_1, \ldots, I_6$ are those given in Example 3.2 and we write $I_j = [a_j, c_j]$ as in the proof of Theorem 3.3.

Acknowledgments

We are grateful to Benjamin Braun and Jonathan Browder for a number of insightful conversations and helpful suggestions. Steven Klee's research was supported by NSF VIGRE Grant DMS-0636297.

References

[Borgs et al. 2006] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, "Counting graph homomorphisms", pp. 315–371 in *Topics in discrete mathematics*, edited by M. Klazar et al., Algorithms Combin. **26**, Springer, Berlin, 2006. MR 2007f:05087 Zbl 1129.05050

[Braun et al. 2011] B. Braun, J. Browder, and S. Klee, "Cellular resolutions of ideals defined by simplicial homomorphisms", preprint, 2011. To appear in *Israel J. Math.* arXiv 1103.1275

[Kozlov 2008] D. Kozlov, *Combinatorial algebraic topology*, Algorithms and Computation in Mathematics **21**, Springer, Berlin, 2008. MR 2008j:55001 Zbl 1130.55001

[Reiner 1997] V. Reiner, "Non-crossing partitions for classical reflection groups", *Discrete Math.* **177**:1-3 (1997), 195–222. MR 99f:06005 Zbl 0892.06001

[Stanley 1997] R. P. Stanley, *Enumerative combinatorics*, vol. 1, Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, Cambridge, 1997. MR 98a:05001 Zbl 0889.05001

Received: 2011-05-27 Accepted: 2011-07-11

lmguerra@ucdavis.edu
Mathematical Sciences Building, One Shields Ave.,
University of California, Davis, CA 95616, United States

klee@math.ucdavis.edu
Mathematical Sciences Building, One Shields Ave.,
University of California, Davis, CA 95616, United States
http://www.math.ucdavis.edu/~klee/

INVOLVE 5:1(2012)

Permutation notations for the exceptional Weyl group F_4

Patricia Cahn, Ruth Haas, Aloysius G. Helminck,
Juan Li and Jeremy Schwartz

(Communicated by Joseph Gallian)

This paper describes a permutation notation for the Weyl groups of type F_4 and G_2. The image in the permutation group is presented as well as an analysis of the structure of the group. This description enables faster computations in these Weyl groups which will prove useful for a variety of applications.

1. Introduction

Weyl groups, or finite Coxeter groups, are widely used in mathematics and in applications (some examples are given in Section 2). They are most commonly represented by generators and relations. The disadvantage of that representation is that elements are not uniquely represented by strings or even minimal strings of generators. For the classical Weyl groups combinatorialists use one-line permutation notation, which corresponds to the orbit of the standard basis vectors under the Weyl group. This combinatorial representation provides unique representation, which makes it efficient for computation (see [Haas and Helminck 2012]). Many properties of the elements, such as length and order, can be quickly read from the combinatorial representation (see, for example, [Haas et al. 2007]). Further, the unique representation provides insight into more complex structures such as involution and twisted involution posets; see [Haas and Helminck 2011].

For the exceptional Weyl groups of type G_2, F_4, E_7 and E_8, the orbit of the standard basis vectors includes not just the positive and negative axes but additional vectors, making description by permutation somewhat less obvious. Nonetheless,

MSC2010: 20G15, 20G20, 22E15, 22E46, 43A85.

Keywords: Weyl groups, Coxeter groups, one-line notation, permutations.

This work was partially supported by NSF grant DMS-0532140 and NSA grant H98230-06-1-0098. Haas was partially supported by NSF grant DMS–0721661, and Helminck by NSF grant DMS-0532140.

similar representations can be made in these cases as well. In this paper we give a permutation representation for the Weyl group of type F_4 and discuss a number of properties of this representation. We also give a similar presentation for G_2.

2. Motivation

Given a field k, symmetric k-varieties are the homogenous spaces G/H, where G is the set of k-rational points of reductive group $\overline{G}$ defined over k and H the set of k-rational points of the set of fixed points of an automorphism σ (defined over k) of the group $\overline{G}$. For k the real or p-adic numbers these are also known as reductive symmetric spaces. These symmetric k-varieties have a detailed fine structure of root systems and Weyl groups, similar to that of the group G itself. This fine structure involves 4 (restricted) root systems and Weyl groups. To study the structure of symmetric k-varieties one needs detailed descriptions of this fine structure and how they act on the various types of elements of these root systems and Weyl groups. For example, to study the representations associated with these symmetric k-varieties one needs a detailed description of the orbits of (minimal) parabolic k-subgroups acting on these symmetric k-varieties. A characterization of these orbits was given in [Helminck and Wang 1993]. They showed that these orbits can be characterized by $\bigcup_{i \in I} W_G(A_i)/W_H(A_i)$, where $\{A_i \mid i \in I\}$ is a set of representatives of the H-conjugacy classes of the σ-stable maximal k-split tori, $W_G(A_i)$ is the set of Weyl group elements that have a representative in $N_G(A_i)$, the normalizer of A_i in G, and $W_H(A_i)$ is the set of Weyl group elements that have a representative in $N_H(A_i)$. To fully classify these orbits one needs to compute the subgroups $W_H(A_i)$ of $W_G(A_i)$. This requires a detailed analysis of the structure of the Weyl groups and their subgroups.

Another example is that the classification of Cartan subspaces can be reduced to a classification of $W_H(A)$-conjugacy classes of σ-singular involutions. The $W_G(A)$-conjugacy classes of involutions were classified in [Helminck 1991]. A detailed analysis of the Weyl groups and their subgroups will enable one to determine how a $W_G(A)$-conjugacy class breaks up in $W_H(A)$-conjugacy classes. There are many other problems related to symmetric k-varieties for which one needs a detailed description of the various Weyl groups and their subgroups. The detailed combinatorial analysis of the structure of the Weyl groups of types F_4 and G_2 in this paper enables us to compute the necessary data to solve those problems for those symmetric k-varieties that have a restricted Weyl group of type F_4 and G_2.

The classical text on Weyl groups is [Bourbaki 2002], while a good modern treatment to Weyl groups and their uses in Lie theory can be found in [Humphreys 1972]. The Weyl groups of type F and G are two of the exceptional Coxeter groups; see [Humphreys 1990] for a basic treatment of these groups.

3. The Weyl group of type F_4

The root system of type F_4 has the following characteristics. There are $n = 48$ roots. The usual basis is the set

$$\left\{\alpha_1 = e_2 - e_3, \alpha_2 = e_3 - e_4, \alpha_3 = e_4, \alpha_4 = \tfrac{1}{2}(e_1 - e_2 - e_3 - e_4)\right\}.$$

The complete set of roots is $\{\pm e_i, \pm e_i \pm e_j, \frac{1}{2}(\pm e_1 \pm e_2 \pm e_3 \pm e_4)\}$. The positive roots are $\{e_i, e_i \pm e_j, \frac{1}{2}(e_1 \pm e_2 \pm e_3 \pm e_4)\}$. Recall the associated Weyl group is generated by the reflections over the hyperplanes orthogonal to the basis roots. These are usually denoted s_{α_i}, which we abbreviate to s_i. Here we label the short positive roots with the numbers 1- 12 and describe how the Weyl group of type F_4 is associated with a subgroup of the permutation group on $[-12, \dots, 12]$. I.e., each element in $W(F_4)$ will be associated with a signed permutation on $\{1, \dots, 12\}$.

To begin compute the images of the short roots under the basis. These are given in Table 1. Each column in the table describes where each root goes under each basis reflection, so when read from top to bottom, the column gives one-line notation for the generators of the Weyl group. These generators are

$$\begin{aligned} s_1 &= (1, 3, 2, 4, 5, 7, 6, 8, 9, 11, 10, 12),\\ s_2 &= (1, 2, 4, 3, 5, 6, 8, 7, 9, 10, 12, 11),\\ s_3 &= (1, 2, 3, -4, 12, 11, 10, 9, 8, 7, 6, 5),\\ s_4 &= (9, 10, 11, 12, -5, 6, 7, 8, 1, 2, 3, 4).\end{aligned}$$

	root r	$s_{\alpha_1}(r)$	$s_{\alpha_2}(r)$	$s_{\alpha_3}(r)$	$s_{\alpha_4}(r)$
1	e_1	1	1	1	9
2	e_2	3	2	2	10
3	e_3	2	4	3	11
4	e_4	4	3	-4	12
5	$\frac{1}{2}(e_1 - e_2 - e_3 - e_4)$	5	5	12	-5
6	$\frac{1}{2}(e_1 - e_2 + e_3 + e_4)$	7	6	11	6
7	$\frac{1}{2}(e_1 + e_2 - e_3 + e_4)$	6	8	10	7
8	$\frac{1}{2}(e_1 + e_2 + e_3 - e_4)$	8	7	9	8
9	$\frac{1}{2}(e_1 + e_2 + e_3 + e_4)$	9	9	8	1
10	$\frac{1}{2}(e_1 + e_2 - e_3 - e_4)$	11	10	7	2
11	$\frac{1}{2}(e_1 - e_2 + e_3 - e_4)$	10	12	6	3
12	$\frac{1}{2}(e_1 - e_2 - e_3 + e_4)$	12	11	5	4

Table 1. Generators of $W(F_4)$.

In cycle notation, they can be expressed as products of transpositions as follows:

$$s_1 = (2,3)(6,7)(10,11), \qquad s_3 = (8,9)(7,10)(6,11)(5,12)(4,-4),$$
$$s_2 = (3,4)(7,8)(11,12), \qquad s_4 = (1,9)(2,10)(3,11)(4,12)(5,-5).$$

Note that the elements of $W(F_4)$ are in one-to-one correspondence with only a subset of signed permutations on $[1,\dots,12]$. In particular, since the first 4 elements give the image of the standard basis of $\mathbb{R}^4$, they determine the other 8 positions uniquely. In Section 3.10 we will see that there are further restrictions on what can occur in the first four places.

3.1. ***A minimal word algorithm.*** We develop a method for the important task of converting from this one-line notation to the standard representation of an element as a minimal word. For $x \in W(F_4)$, recall that the length of x, $l(x)$, is the number of letters in the minimal word of x. It is well-known that the length of x equals the number of positive roots mapped to negative roots by x.

Lemma 3.2. *Any nontrivial element of $W(F_4)$ maps at least one of e_4, $e_2 - e_3$, $e_3 - e_4$, and $\frac{1}{2}(e_1 - e_2 - e_3 - e_4)$ to a negative root.*

Proof. This set of roots is exactly the set of roots which get mapped to negative roots under the basis reflections. □

Lemma 3.3. *Let $x \in W(F_4)$. Then x maps α_i to a negative root if and only if $l(xs_i) < l(x)$.*

Proof. This follows directly from the definitions. □

Algorithm 3.4. Given an element $x = (a_1, a_2, \dots, a_{12}) \in W(F_4)$, the following algorithm will output a minimal word for x.

1. If all $a_i > 0$, go to step 6. Otherwise, go to step 2.
2. If $a_4 < 0$, right multiply by s_3 and go to step 1. Otherwise, go to step 3.
3. If $a_5 < 0$, right multiply by s_4 and go to step 1. Otherwise, go to step 4.
4. If $a_3 < 0$, right multiply by s_2 and go to step 1. Otherwise, go to step 5.
5. Right multiply by s_1 and go to step 1.
6. If the resulting element is not the identity, compare it to the following list in order to determine the final step(s).
   (a) $\{1,3,2,4,5,7,6,8,9,11,10,12\} = s_1$.
   (b) $\{1,2,4,3,5,6,8,7,9,10,12,11\} = s_2$.
   (c) $\{1,3,4,2,5,7,8,6,9,11,12,10\} = s_2 s_1$.
   (d) $\{1,4,2,3,5,8,6,7,9,12,10,11\} = s_1 s_2$.
   (e) $\{1,4,3,2,5,8,7,6,9,12,11,10\} = s_1 s_2 s_1$.

Theorem 3.5. *Algorithm 3.4 produces a minimal word for x.*

Proof. Note that even if $a_i > 0$ for all i, the length of x may not be zero because not all positive roots are represented in the list of twelve roots. In particular, there are six elements of $W(F_4)$ such that $a_i > 0$ for all i. They are precisely those listed in Step 6. of the algorithm together with the identity. Clearly steps 2 and 3 reduce the length of x. If we arrive at step 4, i.e., $a_4 > 0$, and $a_3 < 0$, then one can check that $e_3 - e_4$ maps to a negative root under x, so multiplying by s_2 will reduce the length of x.

If we arrive at step 5, i.e., $a_3, a_4, a_5 > 0$, but some other a_i is negative, then we show that a_2 must be negative. Suppose instead that $a_2 > 0$. Let $\langle i\rangle + \langle j\rangle$ denote the root which is the vector sum of roots i and j. E.g., $\langle 5\rangle = \frac{1}{2}(\langle 1\rangle - \langle 2\rangle - \langle 3\rangle - \langle 4\rangle)$ and since x is a linear map this implies $\langle a_5\rangle = \frac{1}{2}(\langle a_1\rangle - \langle a_2\rangle - \langle a_3\rangle - \langle a_4\rangle)$. Rearranging gives $\langle a_1\rangle = \langle a_2\rangle + \langle a_3\rangle + \langle a_4\rangle + 2\langle a_5\rangle$, with all terms on the right positive by assumption. Therefore, $a_1 > 0$. Similar calculations done in the correct order show that all other a_i must be positive. Explicitly: $\langle a_{10}\rangle = \langle a_5\rangle + \langle a_2\rangle$; $\langle a_{12}\rangle = \langle a_5\rangle + \langle a_4\rangle$; $\langle a_{11}\rangle = \langle a_5\rangle + \langle a_3\rangle$; $\langle a_6\rangle = \langle a_{12}\rangle + \langle a_3\rangle$; $\langle a_7\rangle = \langle a_{12}\rangle + \langle a_2\rangle$; $\langle a_8\rangle = \langle a_{10}\rangle + \langle a_3\rangle$; $\langle a_9\rangle = \frac{1}{2}(\langle a_1\rangle + \langle a_2\rangle + \langle a_3\rangle + \langle a_4\rangle)$.

Thus $a_2 < 0$. In this case $e_2 - e_3$ will be mapped to a negative root, so right multiplication by s_1 will reduce the length.

If we arrive at step 6 then all $a_i > 0$. Clearly these must be products of s_1 and s_2 only. The 5 elements listed above plus the identity are all the possibilities. □

One can determine the length of any $x \in W(F_4)$ by finding a reduced word as above. In what follows we give a combinatorial description of length. Partition the short roots of F_4 into the three sets

$$\alpha = \{\pm 1, \pm 2, \pm 3, \pm 4\}, \quad \beta = \{\pm 5, \pm 6, \pm 7, \pm 8\}, \quad \gamma = \{\pm 9, \pm 10, \pm 11, \pm 12\}.$$

Lemma 3.6. *For all $x \in W(F_4)$, $\{x(\alpha), x(\beta), x(\gamma)\} = \{\alpha, \beta, \gamma\}$. In other words, x permutes the sets α, β and γ.*

Theorem 3.7. *For an element $x = (a_1, a_2, \cdots, a_{12})$, define*

$$N(x) = |\{i : a_i < 0\}|$$

and

$$p(a_i, a_j) = \begin{cases} 0 & \text{if } |a_i| < |a_j| \text{ and } a_i > 0, \\ 2 & \text{if } |a_i| < |a_j| \text{ and } a_i < 0, \\ 1 & \text{if } |a_i| > |a_j|. \end{cases}$$

Find k such that $\{\pm a_{4k+1}, \pm a_{4k+2}, \pm a_{4k+3}, \pm a_{4k+4}\} = \alpha$. If $k = 1$,

$$l(x) = \sum_{i>j} p(a_{4k+i}, a_{4k+j}) + N(x).$$

Otherwise,

$$l(x) = \sum_{i<j} p(a_{4k+i}, a_{4k+j}) + N(x).$$

Proof. The length counts the number of positive roots mapped to negative roots under x. The function $N(x)$ counts the number of short roots mapped to negative roots, while the $p(a_{4k+i}, a_{4k+j})$ terms account for the number of long roots mapped to negative roots. There are three cases depending on which set is mapped to α.

Suppose $x(\alpha) = \alpha$. Each of the positive long roots $e_i \pm e_j$, $i < j$ is the sum or difference of the roots $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$; where the difference is taken as $\langle i\rangle - \langle j\rangle$ where $i < j$. Thus to determine which of these is mapped to a negative long root, we need only consider the sum and difference of $\langle a_i\rangle$ for $i = 1, \dots, 4$. It is easy to check that $\langle a_i\rangle + \langle a_j\rangle$ is a negative root exactly when either $|a_i| > |a_j|$ and $a_j < 0$ or when $|a_i| < |a_j|$ and $a_i < 0$. As well, $\langle a_i\rangle - \langle a_j\rangle$ is negative exactly when $|a_i| < |a_j|$ and $a_i < 0$ or when $|a_i| > |a_j|$ and $a_j > 0$.

Suppose $x(\beta) = \alpha$. Each of the positive long roots $e_i \pm e_j$, $j < i$ is the sum or difference of the roots $\langle 5\rangle$, $\langle 6\rangle$, $\langle 7\rangle$, $\langle 8\rangle$ where the difference is taken as $\langle i\rangle - \langle j\rangle$ where $j < i$. With this reversed order the same conditions for when $\langle a_i\rangle + \langle a_j\rangle$ and $\langle a_i\rangle - \langle a_j\rangle$ are negative will still hold.

Suppose $x(\gamma) = \alpha$. Each of the positive long roots $e_i \pm e_j$, $i < j$ is the sum or difference of the roots $\langle 9\rangle$, $\langle 10\rangle$, $\langle 11\rangle$, $\langle 12\rangle$; where the difference is taken as $\langle i\rangle - \langle j\rangle$ where $i < j$. Again the same conditions hold. □

3.8. ***Group structure and notation properties.*** It is also useful to consider the images of the three sets of short roots as permutations in signed S_4. Refer to the elements in positions 1-4 as set A, the elements in positions 5-8 as set B, and the elements in positions 9-12 as set C. Formally, let $f \in F_4$ such that $f = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12})$. For $1 \le i \le 4$ let $a_i = f_i$ (mod 4), using the representatives $\{1, 2, 3, 4\}$ for $\mathbb{Z}_4$. Similarly, $b_i = f_{4+i}$(mod 4), and $c_i = f_{8+i}$ (mod 4), for $1 \le i \le 4$ again using the representatives $\{1, 2, 3, 4\}$ for $\mathbb{Z}_4$. Let $A = (a_1, a_2, a_3, a_4)$, $B = (b_1, b_2, b_3, b_4)$, and $C = (c_1, c_2, c_3, c_4)$. Denote $(|a_1|, |a_2|, |a_3|, |a_4|)$ by $|A|$, and define $|B|$ and $|C|$ analogously. For example, if $f = (6, -8, 5, -7, 9, 11, -10, 12, -2, 4, 1, 3)$, then $|A| = (2, 4, 1, 3)$, $|B| = (1, 3, 2, 4)$, and $|C| = (2, 4, 1, 3)$.

Theorem 3.9. *The parity of the negations in each block, given the order of the sets* α, β, *and* γ, *is the following*:

set order	block A (1–4)	block B (5–8)	block C (9–12)
$\alpha\beta\gamma$	even	even	even
$\alpha\gamma\beta$	odd	even	even
$\beta\alpha\gamma$	odd	odd	odd
$\beta\gamma\alpha$	even	odd	odd
$\gamma\alpha\beta$	odd	even	odd
$\gamma\beta\alpha$	even	odd	even

Proof. Note that generators s_1 and s_2 do not change the parity of negations in any set, nor do they change the order of the sets. Therefore it suffices to inductively show that this table holds after operating by generators s_3 and s_4 on the right. It is simple to compute using the following rules. s_3 swaps the second and third blocks, and adds or subtracts one negative from the first block. s_4 swaps the first and third blocks, and adds or subtracts one negative from the second block. □

3.10. ***Restrictions on the values of*** $|a_i|$**.** Let $\mathbb{V}$ be the subset of S_4 generated by $(12)(34)$ and $(13)(24)$, and K be the subset of S_4 generated by (23) and (34).

For $X \in S_4$ define $v(X)$ to be the unique element of $\mathbb{V}$ in the coset KX.

Theorem 3.11. *Let $f \in W(F_4)$ with sets A_f, B_f and C_f as defined above. Then $|C_f| = |B_f|v(|A_f|)$, $|B_f| = |A_f|v(|C_f|)$, and $|A_f| = |C_f|v(|B_f|)$.*
Alternative statement: Let $f \in W(F_4)$ with sets f_α, f_β, and f_γ as defined above. Then $f_\gamma = f_\beta v(f_\alpha)$, $f_\beta = f_\alpha v(f_\gamma)$, and $f_\alpha = f_\gamma v(f_\beta)$.

Proof. We proceed by induction. The statement is true for $f = identity$. Assume its true for f, we show its true for $s_i f$ for each s_i. Note that $v((s_i f)_\mu) = v(f_\mu)$ when $i = 1, 2, 4$ and $\mu = \alpha, \beta, \gamma$; $v((s_3 f)_\alpha) = v(f_\alpha)$, $v((s_3 f)_\beta) = (14)(23)v(f_\gamma)$ and $v((s_3 f)_\gamma) = (14)(23)v(f_\beta)$. The cases for s_i where $i \neq 3$ are straightforward.

Furthermore, $(s_3 f)_\beta = (14)(23)f_\gamma$ and $(s_3 f)_\gamma = (14)(23)f_\beta$. These equations provide all of the required components for the proof. For example assume $f_\beta = f_\gamma v(f_\alpha)$. Since $v((s_3 f)_\alpha) = v(f_\alpha)$ and $(s_3 f)_\gamma = (14)(23)f_\beta$ and $(s_3 f)_\beta = (14)(23)f_\gamma$ it follows that $(s_3 f)_\gamma = (s_3 f)_\beta v((s_3 f)_\alpha)$. □

3.12. ***W*****(*****F*****$_4$)** ***as a semidirect product.*** Let F_D denote the subgroup of $W(F_4)$ containing all $d \in F_D$ where $d(\alpha) = \alpha$, $d(\beta) = \beta$, and $d(\gamma) = \gamma$, and let F_S be the subgroup of $W(F_4)$ generated by the generators s_3 and s_4. Let T be the group representing the order of the sets α, β and γ. Define $\tau : W(F_4) \mapsto T$ in the obvious way. Note that $\tau(f) = id$ if and only if $f \in F_D$. Now by Theorem 3.9, the sets α, β and γ occur in order $\alpha\beta\gamma$ in the bottom row notation of f if and only if the permutation A_f contains an even number of negative signs.

Lemma 3.13. *F_D is isomorphic to D_4.*

Proof. The map $\psi : F_D \to D_4$ such that $\psi(f) = A_f$ for $f \in F_D$ provides the isomorphism. □

Theorem 3.14. *$W(F_4) = F_D \rtimes F_S$.*

Proof. We can represent $f \in W(F_4)$ by a pair (d, s) where $f = ds$ and s is the unique element of F_S such that $\tau(s) = \tau(f)$. Define $\phi_s : F_D \mapsto F_D$ where $\phi_s(d) = sds^{-1}$ for $d \in F_D$ and $s \in F_S$. One can check that if $f_1 = d_1 s_1$, represented by the pair (d_1, s_1), and $f_2 = d_2 s_2$, represented by the pair (d_2, s_2), then $f_1 f_2 = d_1 \phi_{s_1}(d_2) s_1 s_2$, represented by the pair $(d_1 \cdot \phi_{s_1}(d_2), s_1 \cdot s_2)$. □

One might hope that this semidirect product would provide an efficient notation for computation in $W(F_4)$. A road block to this seems to be finding a combinatorial description of the multiplication.

4. The Weyl group of type G_2

The root system of type G_2 has the following characteristics. There are $n = 12$ roots. The usual basis is the set $\{\alpha_1 = e_1 - e_2, \alpha_2 = -2e_1 + e_2 + e_3\}$. The complete set of roots is $\{\pm(e_i - e_j)\}$, where $i < j$ and $i, j \in \{1, 2, 3\}$, and $\{\pm(2e_i - e_j - e_k)\}$, where $\{i, j, k\} = \{1, 2, 3\}$. The positive roots are $\{\alpha_1, \alpha_2, \alpha_1 + \alpha_2, 2\alpha_1 + \alpha_2, 3\alpha_1 + \alpha_2, 3\alpha_1 + 2\alpha_2\}$. Again we let s_i denote the reflection over the hyperplane orthogonal to α_i. We label the short positive roots $2\alpha_1 + \alpha_2$, $\alpha_1 + \alpha_2$, and α_1, with the numbers 1-3 respectively, and describe how the Weyl group of type G_2 is associated with a subgroup of the permutation group on $[-3, \dots, 3]$. Here are the images of roots 1, 2, and 3 under the generators of $W(G_2)$:

	root r	$s_{\alpha_1}(r)$	$s_{\alpha_2}(r)$
1	$-e_2 + e_3$	2	1
2	$-e_1 + e_3$	1	3
3	$e_1 - e_2$	-3	2

Reading from top to bottom in each column gives one-line notation for the generators, namely $s_1 = (2, 1, -3)$ and $s_2 = (1, 3, 2)$.

As with $W(F_4)$ we can give a simple combinatorial length formula for $W(G_2)$.

Theorem 4.1. *The length of an element $x = (a_1, a_2, a_3)$ in $W(G_2)$ is given by $l(x) = \sum_{i<j} p(a_i, a_j)$ where $p(a_i, a_j)$ is defined as follows*:

$$p(a_i, a_j) = \begin{cases} 0 & \text{if } |a_i| < |a_j| \text{ and } a_i > 0, \\ 2 & \text{if } |a_i| < |a_j| \text{ and } a_i < 0, \\ 1 & \text{if } |a_i| > |a_j|. \end{cases}$$

Proof. The length counts the number of positive roots mapped to negative roots under x. One can check that the positive roots in $W(G_2)$ are of the form $\langle i \rangle \pm \langle j \rangle$ where $i < j$ and $i, j \in \{1, 2, 3\}$. To determine which of $\langle i \rangle \pm \langle j \rangle$ are mapped to negative roots, we need to determine when $\langle a_i \rangle \pm \langle a_j \rangle$ is a negative root. One can check that $\langle a_i \rangle + \langle a_j \rangle$ is negative when $|a_i| < |a_j|$ and $a_i < 0$, or when $|a_i| > |a_j|$ and $a_j < 0$. Similarly $\langle a_i \rangle - \langle a_j \rangle$ is negative when $|a_i| < |a_j|$ and $a_i < 0$, or when $|a_i| > |a_j|$ and $a_j > 0$. □

Acknowledgements

The results presented here were obtained as part of the NCSU summer REU program. We would like to thank NSF and NSA for their support of this program.

References

[Bourbaki 2002] N. Bourbaki, *Lie groups and Lie algebras. Chapters 4–6*, Elements of Mathematics (Berlin), Springer, Berlin, 2002. MR 2003a:17001 Zbl 0983.17001

[Haas and Helminck 2011] R. Haas and A. G. Helminck, "Admissible sequences for twisted involutions in Weyl groups", *Canad. Math. Bull.* **54**:4 (2011), 663–675. MR 2894516 Zbl 05987122

[Haas and Helminck 2012] R. Haas and A. G. Helminck, "Algorithms for twisted involutions in Weyl groups", *Algebra Colloq.* **19**:2 (2012), 263–282.

[Haas et al. 2007] R. Haas, A. G. Helminck, and N. Rizki, "Properties of twisted involutions in signed permutation notation", *J. Combin. Math. Combin. Comput.* **62** (2007), 121–128. MR 2008d:05163 Zbl 1125.20030

[Helminck 1991] A. G. Helminck, "Tori invariant under an involutorial automorphism. I", *Adv. Math.* **85**:1 (1991), 1–38. MR 92a:20047 Zbl 0731.20029

[Helminck and Wang 1993] A. G. Helminck and S. P. Wang, "On rationality properties of involutions of reductive groups", *Adv. Math.* **99**:1 (1993), 26–96. MR 94d:20051 Zbl 0788.22022

[Humphreys 1972] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics **9**, Springer, New York, 1972. MR 48 #2197 Zbl 0254.17004

[Humphreys 1990] J. E. Humphreys, *Reflection groups and Coxeter groups*, Studies in Advanced Mathematics **29**, Cambridge University Press, Cambridge, 1990. MR 92h:20002 Zbl 0725.20028

Received: 2011-06-25 Revised: 2011-08-22 Accepted: 2011-08-28

patricia.cahn@gmail.com
*Department of Mathematics, Dartmouth College,
Hanover, NH 03755, United States*

rhaas@smith.edu
*Department of Mathematics and Statistics, Smith College,
Northampton, MA 01063, United States*

loek@ncsu.edu
*Department of Mathematics, North Carolina State University,
Raleigh, NC 27695, United States*

jl879@cornell.edu
*School of Operations Research and Information Engineering,
Cornell University, Ithaca, NY 14850, United States*

jschwart@math.umd.edu
*Department of Mathematics, University of Maryland,
College Park, MD 20742, United States*

INVOLVE 5:1(2012)

Progress towards counting D_5 quintic fields

Eric Larson and Larry Rolen

(Communicated by Ken Ono)

Let $N(5, D_5, X)$ be the number of quintic number fields whose Galois closure has Galois group D_5 and whose discriminant is bounded by X. By a conjecture of Malle, we expect that $N(5, D_5, X) \sim C \cdot X^{\frac{1}{2}}$ for some constant C. The best upper bound currently known is $N(5, D_5, X) \ll X^{\frac{3}{4}+\varepsilon}$, and we show this could be improved by counting points on a certain variety defined by a norm equation; computer calculations give strong evidence that this number is $\ll X^{\frac{2}{3}}$. Finally, we show how such norm equations can be helpful by reinterpreting an earlier proof of Wong on upper bounds for A_4 quartic fields in terms of a similar norm equation.

1. Introduction and statement of results

Let K be a number field and $G \leq S_n$ a transitive permutation group on n letters. In order to study the distribution of fields with given degree and Galois group, we introduce the following counting function:

$$N(d, G, X) := \#\{\text{degree } d \text{ number fields } K \text{ with } \mathrm{Gal}(K^{\mathrm{gal}}/\mathbb{Q}) \simeq G \text{ and } |D_K| \leq X\}.$$

Here D_K denotes the discriminant of K, counting conjugate fields as one. Our goal is to study this function for $d = 5$ and $G = D_5$. Malle [2002] has conjectured that

$$N(d, G, X) \sim C(G) \cdot X^{a(G)} \cdot \log(X)^{b(G)-1} \tag{1}$$

for some constant $C(G)$ and for explicit constants $a(G)$ and $b(G)$, and this has been proven for all abelian groups G. Although this conjecture seems to be close to the truth on the whole, Klüners [2005] found a counterexample when $G = C_3 \wr C_2$

MSC2010: primary 11R45; secondary 11R29, 14G05.

Keywords: quintic dihedral number fields, Cohen–Lenstra heuristics for $p = 5$.

The authors are grateful for the support of the NSF in funding the Emory 2011 REU. The authors would like to thank our advisor Andrew Yang, as well as Ken Ono for their guidance, useful conversations, improving the quality of exposition of this article, and hosting the REU.

by showing that the conjecture predicts the wrong value for $b(G)$. This conjecture has been modified to explain all known counterexamples in [Turkelli 2008].

We now turn to the study of $N(5, D_5, X)$. By Malle's conjecture, we expect that

$$N(5, D_5, X) \overset{?}{\sim} C \cdot X^{\frac{1}{2}}. \tag{2}$$

This question is closely related to average 5-parts of class numbers of quadratic fields. In general, let l be a prime, D range over fundamental discriminants, and $r_D := \mathrm{rk}_l(\mathrm{Cl}_{\mathbb{Q}(\sqrt{D})})$. Then the heuristics of Cohen–Lenstra predicts that the average of $l^{r_D} - 1$ over all imaginary quadratic fields is 1, and the average of $l^{r_D} - 1$ over all real quadratic fields is l^{-1}.

In fact, one can show using class field theory that the Cohen–Lenstra heuristics imply that Malle's conjecture is true for D_5 quintic fields. Conversely, the best known upper bound for $N(5, D_5, X)$ is proved using the "trivial" bound (see [Klüners 2006])

$$l^{r_D} \le \#\mathrm{Cl}_{\mathbb{Q}(\sqrt{D})} = O(D^{\frac{1}{2}} \log D). \tag{3}$$

This gives $N(5, D_5, X) \ll X^{\frac{3}{4}+\varepsilon}$, and any improved bound would give nontrivial information on average 5-parts of class groups in a similar manner.

In this paper, we consider a method of point counting on varieties to give upper bounds on $N(5, D_5, X)$. Our main result is the following:

Theorem 1.1. *To any quintic number field K with Galois group D_5, there corresponds a triple (A, B, C) with $A, B \in \mathbb{O}_{\mathbb{Q}[\sqrt{5}]}$ and $C \in \mathbb{Z}$, such that*

$$\mathrm{Nm}_{\mathbb{Q}}^{\mathbb{Q}[\sqrt{5}]}(B^2 - 4 \cdot \bar{A} \cdot A^2) = 5 \cdot C^2 \tag{4}$$

and satisfying the following bounds under any archimedean valuation:

$$|A| \ll D_K^{\frac{1}{4}}, \quad |B| \ll D_K^{\frac{3}{8}}, \quad \text{and} \quad |C| \ll D_K^{\frac{3}{4}}. \tag{5}$$

Conversely, the triple (A, B, C) uniquely determines K.

In Section 6, we further provide numerical evidence that $N(5, D_5, X) \ll X^{\frac{2}{3}+\alpha}$ for very small α; in particular the exponent appears to be much lower than $\frac{3}{4}$.

Before we prove Theorem 1.1, we show that earlier results from [Wong 2005] in the case of $G = A_4$ can be handled in a similar fashion. Namely, we give a shorter proof of the following theorem:

Theorem 1.2 (Wong). *To any quartic number field K with Galois group A_4, there corresponds a tuple $(a_2, a_3, a_4, y) \in \mathbb{Z}^4$ such that*

$$(4a_2^2 + 48a_4)^3 = \mathrm{Nm}_{\mathbb{Q}}^{\mathbb{Q}[\sqrt{-3}]}\big(32a_2^3 + 108a_3^2 - 6a_2(4a_2^2 + 48a_4) - 12\sqrt{-3}y\big)$$

and satisfying the following under any archimedean valuation:

$$|a_2| \ll D_K^{\frac{1}{3}},\ |a_3| \ll D_K^{\frac{1}{2}},\ |a_4| \ll D_K^{\frac{2}{3}},\ \textit{and}\ |y| \ll D_K.$$

Conversely, given such a tuple, there corresponds at most one A_4-quartic field. In particular, we have $N(4, A_4, X) \ll X^{\frac{5}{6}+\varepsilon}$.

2. Upper bounds via point counting

Let G be a transitive permutation group. If K is a number field of discriminant D_K and degree n for which $\mathrm{Gal}(K^{\mathrm{gal}}/\mathbb{Q}) \simeq G$, then Minkowski theory implies there is an element $\alpha \in \mathbb{O}_K$ of trace zero with

$$|\alpha| \ll D_K^{\frac{1}{2(n-1)}} \quad \text{(under any archimedean valuation)},$$

where the implied constant depends only on n. In particular, if K is a primitive extension of $\mathbb{Q}$, then $K = \mathbb{Q}(\alpha)$, so the characteristic polynomial of α will determine K. One can use this to give an upper bound on $N(n, G, X)$ (at least in the case where K is primitive), since every pair (K, α) as above gives a $\mathbb{Z}$-point of

$$\operatorname{Spec} \mathbb{Q}[x_1, x_2, \ldots, x_n]^G/(s_1),$$

where $s_1 = x_1 + x_2 + \cdots + x_n$ (here $\mathbb{Q}[x_1, x_2, \ldots, x_n]^G$ denotes the ring of G-invariant polynomials in $\mathbb{Q}[x_1, x_2, \ldots, x_n]$).

3. Proof of Theorem 1.2

In this section, we sketch a simplified (although essentially equivalent) version of Wong's proof [Wong 2005] that $N(4, A_4, X) \ll X^{\frac{5}{6}+\epsilon}$ as motivation for our main theorem. In this section, we assume that the reader is familiar with the arguments in Wong's paper. As noted in the last section, it suffices to count triples (a_2, a_3, a_4) for which $|a_k| \ll X^{\frac{k}{6}}$ under any archimedean valuation and

$$\begin{aligned}256a_4^3 - 128a_2^2a_4^2 + (16a_2^4 + 144a_2a_3^2)a_4 - 4a_2^3a_3^2 - 27a_3^4 \\ = \mathrm{Disc}(x^4 + a_2x^2 + a_3x + a_4) = y^2\end{aligned}$$

for some $y \in \mathbb{Z}$. (See Equation 4.2 of [Wong 2005].)

The key observation of Wong's paper (although he does not state it in this way) is that this equation can be rearranged as

$$(4a_2^2 + 48a_4)^3 = \mathrm{Nm}_{\mathbb{Q}}^{\mathbb{Q}[\sqrt{-3}]}\big(32a_2^3 + 108a_3^2 - 6a_2(4a_2^2 + 48a_4) - 12\sqrt{-3}y\big). \quad (6)$$

One now notes that there are $\ll X^{\frac{2}{3}}$ possibilities for $4a_2^2 + 48a_4$, and for each of these choices $(4a_2^2 + 48a_4)^3$ can be written in $\ll X^{\varepsilon}$ ways as a norm of an element of $\mathbb{Q}[\sqrt{-3}]$. Thus, it suffices to count the number of points (a_2, a_3) for which

$$32a_2^3 + 108a_3^2 - 6a_2(4a_2^2 + 48a_4) - 12\sqrt{-3}y \quad \text{and} \quad 4a_2^2 + 48a_4$$

are fixed. But the above equation defines an elliptic curve, on which the number of integral points can be bounded by Theorem 3 in [Heath-Brown 2002]. This then gives Wong's bound (as well as the conditional bound assuming standard conjectures as Wong shows).

4. Proof of Theorem 1.1

In this section, we give the proof of Theorem 1.1. As explained in Section 2, it suffices to understand the $\mathbb{Z}$-points of

$$\operatorname{Spec} \mathbb{Q}[x_1, x_2, x_3, x_4, x_5]^{D_5}/(x_1 + x_2 + x_3 + x_4 + x_5)$$

inside a particular box. Write ζ for a primitive fifth root of unity, and define

$$V_j = \sum_{i=1}^{5} \zeta^{ij} x_i.$$

In terms of the V_j, we define

$$\begin{aligned} A &= V_2 \cdot V_3, \\ B &= V_1 \cdot V_2^2 + V_3^2 \cdot V_4, \\ C &= \frac{1}{\sqrt{5}} \cdot (V_1 \cdot V_2^2 - V_3^2 \cdot V_4) \cdot (V_2 \cdot V_4^2 - V_1^2 \cdot V_3). \end{aligned}$$

Lemma 4.1. *The expressions A, B, and C are invariant under the action of D_5.*

Proof. The generators of D_5 act by $V_j \mapsto V_{5-j}$ and $V_j \mapsto \zeta^j V_j$; the result follows immediately. □

Lemma 4.2. *We have $A, B \in \mathcal{O}_{\mathbb{Q}[\sqrt{5}]}$ and $C \in \mathbb{Z}$.*

Proof. To see the first assertion, it suffices to show that A and B are invariant by the element of $\operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q})$ given by $\zeta \mapsto \zeta^{-1}$. But this induces the map $V_j \mapsto V_{5-j}$, so this is clear.

To see that C is in $\mathbb{Z}$, we observe that the generator of $\operatorname{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q})$ given by $\zeta \mapsto \zeta^2$ acts by $C\sqrt{5} \mapsto -C\sqrt{5}$. Since $C\sqrt{5}$ is an algebraic integer, it follows that $C\sqrt{5}$ must be a rational integer times $\sqrt{5}$, so $C \in \mathbb{Z}$. □

Now, we compute

$$B^2 - 4 \cdot \bar{A} \cdot A^2 = (V_1 \cdot V_2^2 + V_3^2 \cdot V_4)^2 - 4 \cdot V_1 \cdot V_4 \cdot (V_2 \cdot V_3)^2 = (V_1 \cdot V_2^2 - V_3^2 \cdot V_4)^2.$$

Therefore,

$$\mathrm{Nm}_{\mathbb{Q}}^{\mathbb{Q}[\sqrt{5}]}(B^2 - 4 \cdot \bar{A} \cdot A^2) = (V_1 \cdot V_2^2 - V_3^2 \cdot V_4)^2 \cdot (V_2 \cdot V_4^2 - V_1^2 \cdot V_3)^2 = 5 \cdot C^2,$$

which verifies the identity claimed in Theorem 1.1.

To finish the proof, it remains to show that to each triple (A, B, C), there corresponds at most one D_5-quintic field. To do this, we begin with the following lemma.

Lemma 4.3. *None of the V_j are zero.*

Proof. Suppose that some V_j is zero. Since $\bar{A} \cdot A^2 = V_1 \cdot V_2^2 \cdot V_3^2 \cdot V_4$, it follows that $\bar{A} \cdot A^2 = 0$, and hence

$$\mathrm{Nm}_{\mathbb{Q}}^{\mathbb{Q}[\sqrt{5}]}(B^2) = 5 \cdot C^2,$$

which implies $B = C = 0$. Using $B = 0$, we have $V_1 V_2^2 \cdot V_3^2 V_4 = V_1 V_2^2 + V_3^2 V_4 = 0$, so $V_1 V_2^2 = V_3^2 V_4 = 0$. Similarly, using $\bar{B} = 0$, we have $V_2 V_4^2 = V_1^2 V_3 = 0$. Thus, all pairwise products $V_i V_j$ with $i \neq j$ are zero, so at most one V_k is nonzero. Solving for the x_i, we find $x_i = \zeta^{-ik} c$ for some constant c. (It is easy to verify that this is a solution, since $\sum \zeta^i = 0$; it is unique up to rescaling because the transformation $(x_i) \mapsto (V_i)$ is given by a Vandermonde matrix of rank 4). Hence, the minimal polynomial of α is $t^5 - c^5 = 0$, which is visibly not a D_5 extension. □

Lemma 4.4. *For fixed (A, B, C), there are at most two possibilities for the ordered quadruple*

$$(V_1 V_2^2, V_3^2 V_4, V_2 V_4^2, V_1^2 V_3).$$

Proof. Since $V_1 V_2^2 + V_3^2 V_4 = B$ and $V_1 V_2^2 \cdot V_3^2 V_4 = \bar{A} \cdot A^2$ are determined, there are at most two possibilities for the ordered pair $(V_1 V_2^2, V_3^2 V_4)$. Similarly, there at most two possibilities for the ordered pair $(V_2 V_4^2, V_1^2 V_3)$; thus if $V_1 V_2^2 = V_3^2 V_4$, then we are done. Otherwise,

$$V_2 \cdot V_4^2 - V_1^2 \cdot V_3 = \frac{C\sqrt{5}}{V_1 \cdot V_2^2 - V_3^2 \cdot V_4}.$$

Since $V_2 V_4^2 + V_1^2 V_3 = \bar{B}$, this shows that the ordered pair $(V_1 V_2^2, V_3^2 V_4)$ determines $(V_2 V_4^2, V_1^2 V_3)$. Hence there are at most two possibilities our ordered quadruple. □

Lemma 4.5. *For fixed (A, B, C), there are at most ten possibilities for the ordered quadruple (V_1, V_2, V_3, V_4).*

Proof. In light of Lemmas 4.4 and 4.3, it suffices to show there at most five possibilities for (V_1, V_2, V_3, V_4) when we have fixed nonzero values for

$$(V_1 V_4, V_2 V_3, V_1 V_2^2, V_3^2 V_4, V_2 V_4^2, V_1^2 V_3).$$

But this follows from the identities

$$V_1^5 = \frac{V_1V_2^2\cdot(V_1^2V_3)^2}{(V_2V_3)^2}, \quad V_3 = \frac{V_1^2V_3}{V_1^2}, \quad V_4 = \frac{V_3^2V_4}{V_3^2}, \quad V_2 = \frac{V_2V_4^2}{V_4^2}. \quad \square$$

This completes the proof of Theorem 1.1, because $|D_5| = 10$, so each D_5-quintic field corresponds to ten ordered quadruples (V_1, V_2, V_3, V_4), each of which can be seen to correspond to the same triple (A, B, C). Thus, the triple (A, B, C) uniquely determines the D_5-quintic field, since otherwise we would have at least 20 quadruples (V_1, V_2, V_3, V_4) corresponding to (A, B, C), contradicting Lemma 4.5.

5. The quadratic subfield

Proposition 5.1. *Suppose that K is a D_5-quintic field corresponding to a triple (A, B, C) with $C \neq 0$. Then the composite of $\mathbb{Q}[\sqrt{5}]$ with the unique quadratic subfield $F \subset K^{gal}$ is generated by adjoining to $\mathbb{Q}[\sqrt{5}]$ the square root of*

$$(2\sqrt{5} - 10)\cdot(B^2 - 4\cdot\bar{A}\cdot A^2).$$

Proof. Using the results of the previous section, we note that

$$\sqrt{(2\sqrt{5}-10)\cdot(B^2-4\cdot\bar{A}\cdot A^2)} = 2\cdot(\zeta-\zeta^{-1})\cdot(V_1\cdot V_2^2 - V_3^2\cdot V_4).$$

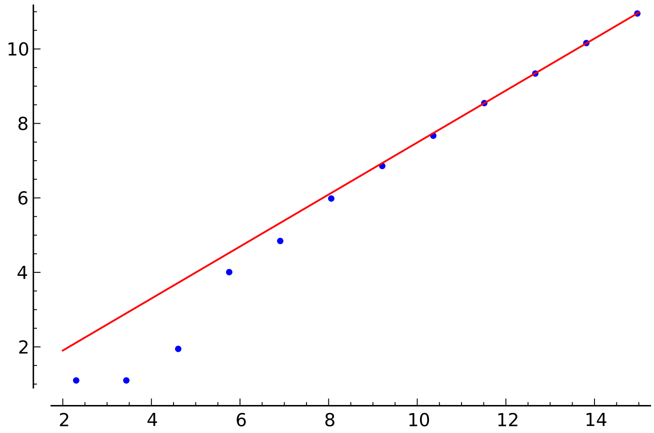
By inspection, the D_5-action on the above expression is by the sign representation, and the action of $\mathrm{Gal}(\mathbb{Q}[\zeta]/\mathbb{Q}[\sqrt{5}])$ is trivial. Hence, adjoining the above quantity to $\mathbb{Q}[\sqrt{5}]$ generates the composite of $\mathbb{Q}[\sqrt{5}]$ with the quadratic subfield F. □

6. Discussion of computational results

Numerical evidence indicates that the number of triples (A, B, C) satisfying the conditions of Theorem 1.1 is $O(X^{\frac{2}{3}+\alpha})$ for a small number α (in particular, much less than $O(X^{\frac{3}{4}})$). More precisely, we have the following table of results. The computation took approximately four hours on a 3.3 GHz CPU, using the program available at http://web.mit.edu/~elarson3/www/d5-count.py.

X	$\#(A,B,C)$	X	$\#(A,B,C)$	X	$\#(A,B,C)$
10	3	1000	127	100000	5145
31	3	3162	397	316227	11385
100	7	10000	951	1000000	25807
316	55	31622	2143	3162277	57079

The log plot on the next page shows that after the first few data points, the least squares best fit to the last four data points given by $y = 0.698x + 0.506$ with slope a little more than $\frac{2}{3}$ is quite close.



References

[Heath-Brown 2002] D. R. Heath-Brown, "The density of rational points on curves and surfaces", *Ann. of Math.* (2) **155**:2 (2002), 553–595. MR 2003d:11091 Zbl 1039.11044

[Klüners 2005] J. Klüners, "A counterexample to Malle's conjecture on the asymptotics of discriminants", *C. R. Math. Acad. Sci. Paris* **340**:6 (2005), 411–414. MR 2005m:11214 Zbl 1083.11069

[Klüners 2006] J. Klüners, "Asymptotics of number fields and the Cohen–Lenstra heuristics", *J. Théor. Nombres Bordeaux* **18**:3 (2006), 607–615. MR 2008j:11162 Zbl 1142.11078

[Malle 2002] G. Malle, "On the distribution of Galois groups", *J. Number Theory* **92**:2 (2002), 315–329. MR 2002k:12010 Zbl 1022.11058

[Turkelli 2008] S. Turkelli, "Connected components of Hurwitz schemes and Malle's conjecture", preprint, 2008. arXiv 0809.0951

[Wong 2005] S. Wong, "Densities of quartic fields with even Galois groups", *Proc. Amer. Math. Soc.* **133**:10 (2005), 2873–2881. MR 2006d:11138 Zbl 1106.11041

Received: 2011-07-20 Accepted: 2011-08-04

elarson3@gmail.com
Department of Mathematics, Harvard,
Cambridge, MA 02138, United States

larry.rolen@mathcs.emory.edu
Department of Mathematics and Computer Science,
Emory University, Atlanta, GA 30322, United States

INVOLVE 5:1(2012)

On supersingular elliptic curves and hypergeometric functions

Keenan Monks

(Communicated by Ken Ono)

The Legendre family of elliptic curves has the remarkable property that both its periods and its supersingular locus have descriptions in terms of the hypergeometric function ${}_2F_1\left(\begin{smallmatrix}1/2 & 1/2\\ & 1\end{smallmatrix} \;\middle|\; z\right)$. In this work we study elliptic curves and elliptic integrals with respect to the hypergeometric functions ${}_2F_1\left(\begin{smallmatrix}1/3 & 2/3\\ & 1\end{smallmatrix} \;\middle|\; z\right)$ and ${}_2F_1\left(\begin{smallmatrix}1/2 & 5/12\\ & 1\end{smallmatrix} \;\middle|\; z\right)$, and prove that the supersingular λ-invariant locus of certain families of elliptic curves are given by these functions.

1. Introduction and statement of results

Let p be a prime and $\mathbb{F}$ a field of characteristic p. An *elliptic curve* $E/\mathbb{F}$ is a curve of the form

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

where $a_i \in \mathbb{F}$ and the points in E are elements of $\overline{\mathbb{F}} \times \overline{\mathbb{F}}$. This curve must be nonsingular in that it has no multiple roots. A point at infinity must also be included on the curve to make it projective.

There is an important invariant defined for any isomorphism class of elliptic curves (two curves are isomorphic if they have the same defining equation up to some change of coordinate system). Using the notation of an elliptic curve as before, the j-invariant $j(E)$ and discriminant $\Delta(E)$ are defined to be

$$j(E) = \frac{c_4^3}{\Delta}$$

and

$$\Delta(E) = \frac{c_4^3 - c_6^2}{1728}$$

where $c_4 = b_2^2 - 24b_4$, $c_6 = -b_2^3 + 36b_2b_4 - 216a_3^2 - 864a_6$, $b_2 = a_1^2 + 4a_2$, and $b_4 = a_1a_3 + 2a_4$.

MSC2010: 14H52, 33C05.

Keywords: elliptic curves, hypergeometric functions.

It is well-known that the points on the curve E with coordinates in $\overline{\mathbb{F}}$ form the group $E(\mathbb{F})$ (see [Washington 2003] for an explanation of the group structure). The curve E is called *supersingular* if and only if the group $E(\mathbb{F})$ has no p-torsion. In this paper, we will determine when certain infinite families of elliptic curves are supersingular for any prime.

One well-known and widely studied family of elliptic curves is the Legendre family, which we denote by

$$E_{\frac{1}{2}}(\lambda) : y^2 = x(x-1)(x-\lambda)$$

for $\lambda \neq 0, 1$. We define its *supersingular locus* by

$$S_{p,\frac{1}{2}}(\lambda) := \prod_{\substack{\lambda_0 \in \overline{\mathbb{F}}_p \\ \text{supersingular} E_{\frac{1}{2}}(\lambda_0)}} (\lambda - \lambda_0).$$

The locus $S_{p,\frac{1}{2}}(\lambda)$ and the periods of $E_{\frac{1}{2}}(\lambda)$ have beautiful and simple descriptions in terms of the hypergeometric function

$${}_2F_1\left(\begin{matrix} a \ b \\ c \end{matrix} \,\middle|\, z\right) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}.$$

Here $a, b, z \in \mathbb{C}$, $c \in \mathbb{C} \setminus \mathbb{Z}^{\leq 0}$, $(x)_0 = 1$, and $(x)_n = (x)(x+1)\cdots(x+n-1)$ is the Pochhammer symbol. For any prime p, define

$${}_2F_1\left(\begin{matrix} a \ b \\ c \end{matrix} \,\middle|\, z\right)_p \equiv \sum_{n=0}^{p-1} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} \pmod{p}.$$

It is natural to study hypergeometric functions related to elliptic integrals. An *elliptic integral of the first kind* is written as

$$K(k) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1 - k^2 \sin^2(\theta)}}.$$

From [Borwein and Borwein 1987] we have the following identities for appropriate ranges of k:

$$K(k) = \frac{\pi}{2}\, {}_2F_1\left(\begin{matrix} \frac{1}{2} \ \frac{1}{2} \\ 1 \end{matrix} \,\middle|\, k^2\right), \tag{1-1a}$$

$$K^2(k) = \frac{\pi^2}{4} \sqrt{\frac{1 - \frac{8}{9}h^2}{1 - (kk')^2}} \left({}_2F_1\left(\begin{matrix} \frac{1}{3} \ \frac{2}{3} \\ 1 \end{matrix} \,\middle|\, h^2\right)\right)^2, \tag{1-1b}$$

$$K(k) = \frac{\pi}{2}(1 - (2kk')^2)^{-\frac{1}{4}}\, {}_2F_1\left(\begin{matrix} \frac{1}{4} \ \frac{3}{4} \\ 1 \end{matrix} \,\middle|\, \frac{(2kk')^2}{(2kk')^2 - 1}\right), \tag{1-1c}$$

$$K(k) = \frac{\pi}{2}(1-(kk')^2)^{-\frac{1}{4}}\, {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix} \,\middle|\, J^{-1}\right). \tag{1-1d}$$

Here $k' = \sqrt{1-k^2}$, $J = \dfrac{(4(2kk')^{-2}-1)^3}{27(2kk')^{-2}}$ and h is the smaller of the two solutions of

$$\frac{(9-8h^2)^3}{64h^6h'^2} = J.$$

For the locus $S_{p,\frac{1}{2}}$, it is a classical result (see [Husemöller 2004] and [Silverman 1986]) that

$$S_{p,\frac{1}{2}}(\lambda) \equiv {}_2F_1\left(\begin{matrix}\frac{1}{2} & \frac{1}{2}\\ & 1\end{matrix} \,\middle|\, \lambda\right)_p \pmod p.$$

In [El-Guindy and Ono 2012], El-Guindy and Ono studied the family of curves defined by

$$E_{\frac{1}{4}}(\lambda) : y^2 = (x-1)(x^2+\lambda).$$

They proved a result analogous to the classical case, namely

$$\prod_{\substack{\lambda_0\in\overline{\mathbb{F}}_p\\ \text{supersingular } E_{\frac{1}{4}}(\lambda_0)}} (\lambda-\lambda_0) \equiv {}_2F_1\left(\begin{matrix}\frac{1}{4} & \frac{3}{4}\\ & 1\end{matrix} \,\middle|\, -\lambda\right)_p \pmod p.$$

Here we prove two other cases of this phenomenon that cover the other hypergeometric functions related to elliptic integrals listed in (1-1). We define the following families of elliptic curves:

$$E_{\frac{1}{3}}(\lambda) : y^2 + \lambda yx + \lambda^2 y = x^3, \tag{1-2}$$

$$E_{\frac{1}{12}}(\lambda) : y^2 = 4x^3 - 27\lambda x - 27\lambda. \tag{1-3}$$

We note that $E_{\frac{1}{3}}(\lambda)$ is singular for $\lambda \in \{0, 27\}$, and that $E_{\frac{1}{12}}(\lambda)$ is singular for $\lambda \in \{0, 1\}$.

We also define, for each $i \in \left\{\frac{1}{3}, \frac{1}{4}, \frac{1}{12}\right\}$ and all primes $p \geq 5$,

$$S_{p,i}(\lambda) := \prod_{\substack{\lambda_0\in\overline{\mathbb{F}}_p\\ \text{supersingular } E_i(\lambda_0)}} (\lambda-\lambda_0).$$

Generalizing the results above, we prove the following for $E_{\frac{1}{3}}(\lambda)$ and $E_{\frac{1}{12}}(\lambda)$.

Theorem 1.1. *For any prime $p \geq 5$, we have*

$$S_{p,\frac{1}{3}}(\lambda) \equiv \lambda^{\lfloor \frac{p}{3} \rfloor}\, {}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix} \,\middle|\, \frac{27}{\lambda}\right)_p \pmod p.$$

Theorem 1.2. *For any prime $p \geq 5$, we have the following*:

(1) *If* $p \equiv 1, 5 \pmod{12}$, *then*

$$S_{p,\frac{1}{12}}(\lambda) \equiv {c_p}^{-1} \lambda^{\lfloor \frac{p}{12} \rfloor} {}_2F_1\left(\begin{matrix} \frac{1}{12} & \frac{5}{12} \\ & 1 \end{matrix} \,\middle|\, 1 - \frac{1}{\lambda}\right)_p \pmod{p},$$

(2) *if* $p \equiv 7, 11 \pmod{12}$, *then*

$$S_{p,\frac{1}{12}}(\lambda) \equiv {c_p}^{-1} \lambda^{\lfloor \frac{p}{12} \rfloor} {}_2F_1\left(\begin{matrix} \frac{7}{12} & \frac{11}{12} \\ & 1 \end{matrix} \,\middle|\, 1 - \frac{1}{\lambda}\right)_p \pmod{p},$$

where $c_p = \binom{6 \lfloor \frac{p}{12} \rfloor + d_p}{\lfloor \frac{p}{12} \rfloor}$, *and* $d_p = 0, 2, 2, 4$ *for* $p \equiv 1, 5, 7, 11 \pmod{12}$.

Remark. The j-invariant of $E_{\frac{1}{3}}(\lambda)$ is $\lambda(\lambda - 24)^3/(\lambda - 27)$ and the j-invariant of $E_{\frac{1}{12}}(\lambda)$ is $1728\lambda/(\lambda - 1)$. Notice that $E_{\frac{1}{3}}(\lambda)$ is singular when $\lambda = 0$ and $j = 0$. Also, $E_{\frac{1}{12}}(\lambda)$ is singular when its j-invariant is 0 and undefined when $j = 1728$.

In addition to the stated result, the proof of Theorem 1.2 yields some fascinating combinatorial identities as well. The following is one such identity obtained for a specific class of p modulo 12. Similar results also hold for primes in the other congruence classes, but are omitted for brevity.

Corollary 1.3. *Let* $p \geq 5$ *be a prime congruent to* 1 *modulo* 12, *and let* $m = \frac{p-1}{12}$. *Then for all* $0 \leq n \leq m$,

$$4^n \binom{3m - n}{3m - 3n} \binom{6m}{3m - n} \binom{6m}{m} \equiv 27^n \sum_{t=n}^{m} \binom{m}{t} \binom{5m}{t} \binom{6m}{3m} \pmod{p}.$$

In particular, *when* $n = m$,

$$4^m \binom{6m}{2m} \binom{6m}{m} \equiv 27^m \binom{5m}{m} \binom{6m}{3m} \pmod{p}.$$

2. Preliminaries

Throughout, let $p \geq 5$ be prime.

Definition 2.1. The *Hasse invariant* of an elliptic curve defined by $f(w, x, y) = 0$ is the coefficient of $(wxy)^{p-1}$ in $f(w, x, y)^{p-1}$. Likewise, the *Hasse invariant* of a curve defined by $y^2 = f(x)$ is the coefficient of x^{p-1} in $f(x)^{\frac{p-1}{2}}$.

Remark. The projective completions of $E_{\frac{1}{3}}(\lambda)$ and $E_{\frac{1}{12}}(\lambda)$ are

$$wy^2 + \lambda wxy + \lambda^2 y - x^3 = 0$$

and

$$wy^2 - 4x^3 + 27\lambda w^2 x + 27\lambda w^3 = 0.$$

Here is a well-known characterization of supersingular elliptic curves.

Lemma 2.2 [Husemöller 2004, Definition 3.1 of Chapter 13]. *An elliptic curve E is supersingular if and only if its Hasse invariant is* 0.

It is well-known that two elliptic curves defined over $\overline{\mathbb{F}}_p$ are isomorphic if and only if they have the same j-invariant. Recall the following formula for the number of isomorphism classes of supersingular elliptic curves over $\overline{\mathbb{F}}_p$ (see [Washington 2003]). We write $p-1=12m_p+6\epsilon_p+4\delta_p$, where $\epsilon_p, \delta_p \in \{0, 1\}$.

Lemma 2.3. *Up to isomorphism, there are exactly*

$$m_p+\epsilon_p+\delta_p$$

supersingular elliptic curves in characteristic p.

Remark. It is known that $\delta_p = 1$ only when $p \equiv 2 \pmod 3$ (i.e., when 0 is a supersingular j-invariant) and $\epsilon_p = 1$ only when $p \equiv 3 \pmod 4$ (when 1728 is a supersingular j-invariant). Also, in all cases $m_p = \left\lfloor \frac{p}{12} \right\rfloor$.

3. Proof of main results

We first prove several preliminary lemmas.

Lemma 3.1. *There are exactly* $\left\lfloor \frac{p}{3} \right\rfloor$ *distinct values of* λ *for which* $E_{\frac{1}{3}}(\lambda)$ *is supersingular over* $\overline{\mathbb{F}}_p$.

Proof. To calculate the degree of $S_{p,\frac{1}{3}}(\lambda)$, we must consider how many different values for λ yield a curve $E_{\frac{1}{3}}(\lambda)$ with a given supersingular j-invariant. From [Lennon 2010] we have that

$$j(E_{\frac{1}{3}}(\lambda)) = \frac{\lambda(\lambda-24)^3}{\lambda-27} \tag{3-1}$$

and that the discriminant $\Delta(E_{\frac{1}{3}}(\lambda)) = \lambda^8(\lambda-27)$. Hence there are usually four λ-invariants for a given j-invariant, but there are certain exceptions. Since the only roots of Δ in this case are 0 and 27, we know that these and 1728 are the only possible j-invariants for which there are less than four corresponding λ-invariants. However, there are four distinct values of λ for which $j(E_{\frac{1}{3}}(\lambda)) = 27$. Also, only $\lambda = 18 \pm 6\sqrt{3}$ gives a value of 1728 for j, so the correspondence is 2-to-1 in this case. As mentioned previously, the curve is singular for $\lambda = 0$, so the only value of λ that will give a j-invariant of 0 is $\lambda = 24$. The correspondence is thus one-to-one for $j = 0$.

Using the ideas of Lemma 2.3, we have that each of the m_p supersingular j-invariants is obtained from four supersingular λ-invariants, δ_p can come from at most one λ-invariant, and ϵ_p comes from two, if any, λ-invariants. Thus the total number of λ-invariants, and the degree of $S_{p,\frac{1}{3}}(\lambda)$, is

$$4m_p+\delta_p+2\epsilon_p=4\left\lfloor\frac{p}{12}\right\rfloor+\delta_p+2\epsilon_p.$$

It is easily verified that this equals $\left\lfloor\frac{p}{3}\right\rfloor$ for every prime p, and so we are done. □

Lemma 3.2. *There are exactly $\left\lfloor\frac{p}{12}\right\rfloor$ distinct values of λ for which $E_{\frac{1}{12}}(\lambda)$ is supersingular over $\overline{\mathbb{F}}_p$.*

Proof. The j-invariant of $E_{\frac{1}{12}}(\lambda)$ is

$$j(E_{\frac{1}{12}}(\lambda))=\frac{1728\lambda}{\lambda-1}. \tag{3-2}$$

This is a one-to-one correspondence from λ-invariants to j-invariants for $j\neq 1728$. Also, the special cases $j=0$ and $j=1728$ do not apply here, for the curve is singular for these respective j-invariants. Thus by Lemma 2.3 there are exactly $\left\lfloor\frac{p}{12}\right\rfloor$ values of λ for which $E_{\frac{1}{12}}(\lambda)$ is supersingular. □

Proof of Theorem 1.1. The curve $E_{\frac{1}{3}}(\lambda)$ can be defined as

$$f(w,x,y)=wy^2+\lambda wxy+\lambda^2w^2y-x^3=0.$$

To compute its Hasse invariant, we consider a general term in the expansion of $(wy^2+\lambda wxy+\lambda^2w^2y-x^3)^{p-1}$. It has the form

$$(wy^2)^a(\lambda wxy)^b(\lambda^2w^2y)^c(-x^3)^d,$$

where $a+b+c+d=p-1$. In order for this to be a constant multiple of a power of wxy, we must have $a=c=d$.

Thus the terms that we are concerned with are of the form

$$(wy^2)^n(\lambda^2w^2y)^n(-x^3)^n(\lambda wxy)^{p-3n-1}=(-\lambda)^{p-n-1}(wxy)^{p-1}.$$

For a given n, there are

$$\binom{p-1}{n}\binom{p-n-1}{n}\binom{p-2n-1}{n}$$

ways to choose which of the $f(w,x,y)$ factors we obtain each of the wy^2, λ^2w^2y, and $-x^3$ terms from. Summing over all possible values of n, we determine the

Hasse invariant to be

$$\begin{aligned}
&\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\binom{p-1}{n}\binom{p-n-1}{n}\binom{p-2n-1}{n}(-\lambda)^{p-n-1}\\
&\equiv\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\frac{(-\lambda)^{p-n-1}(p-1)(p-2)\cdots(p-n)}{n!}\\
&\qquad\cdot\frac{(p-n-1)\cdots(p-2n)}{n!}\\
&\qquad\cdot\frac{(p-2n-1)\cdots(p-3n)}{n!}\pmod{p}\\
&\equiv\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\frac{(3n)!}{n!^3}\lambda^{p-n-1}\pmod{p}.
\end{aligned}$$

By definition, we have

$${}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p\equiv\sum_{n=0}^{p-1}\frac{\left(\frac{1}{3}\right)_n\left(\frac{2}{3}\right)_n}{n!^2}\frac{27^n}{x^n}\pmod{p}.$$

However, if $n>\lfloor\frac{p}{3}\rfloor$, then p will appear in the numerator of either $\left(\frac{1}{3}\right)_n$ or $\left(\frac{2}{3}\right)_n$, making those terms congruent to 0 modulo p, so

$$\begin{aligned}
\lambda^{p-1}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p&\equiv\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\frac{\left(\frac{1}{3}\right)_n\left(\frac{2}{3}\right)_n}{n!^2}27^n\lambda^{p-n-1}\pmod{p}\\
&\equiv\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\frac{27^n\frac{1}{3}\frac{2}{3}\frac{4}{3}\frac{5}{3}\cdots\frac{3n-2}{3}\frac{3n-1}{3}}{n!^2}\lambda^{p-n-1}\pmod{p}\\
&\equiv\sum_{n=0}^{\lfloor \frac{p}{3}\rfloor}\frac{(3n)!}{n!^3}\lambda^{p-n-1}\pmod{p}.
\end{aligned}$$

Thus $\lambda^{p-1}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ is congruent modulo p to the Hasse invariant of $E_{\frac{1}{3}}(\lambda)$. So by Lemma 2.2, λ is a root of $\lambda^{p-1}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p\equiv 0\pmod{p}$ if and only if $E_{\frac{1}{3}}(\lambda)$ is supersingular, i.e., if and only if λ is a root of $S_{p,\frac{1}{3}}(x)$.

The least power of λ in ${}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ is $-\lfloor\frac{p}{3}\rfloor$. Hence $\lambda^{\lfloor p/3\rfloor}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ has the same roots as $\lambda^{p-1}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$, with the exception of 0, which is not a λ-invariant as shown in Lemma 3.1, and thus is not a root of $S_{p,\frac{1}{3}}$.

The degree of $\lambda^{\lfloor \frac{p}{3} \rfloor}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ is exactly $\lfloor \frac{p}{3} \rfloor$. Since the degree of $S_{p,\frac{1}{3}}(\lambda)$ is also $\lfloor \frac{p}{3} \rfloor$ by Lemma 3.1, it follows that $\lambda^{\lfloor \frac{p}{3} \rfloor}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p \equiv c \cdot S_{p,\frac{1}{3}}(\lambda)$ (mod p). However, c is 1 since $\lambda^{\lfloor \frac{p}{3} \rfloor}\,{}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ is monic: we are done. □

Proof of Theorem 1.2. Assume $p \equiv 1, 5 \pmod{12}$. The function

$$f(z) = {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\,z\right)$$

satisfies the second order differential equation

$$z(1-z)\frac{d^2f}{dz^2} + \left(1 - \frac{3}{2}z\right)\frac{df}{dz} - \frac{5}{144}f = 0.$$

Substituting $z = 1 - \frac{1}{x}$, we see that $g(x) = {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\,1-\frac{1}{x}\right)$ satisfies

$$x^2(x-1)\frac{d^2g}{dx^2} + x\left(\frac{3}{2}x - \frac{1}{2}\right)\frac{dg}{dx} - \frac{5}{144}g = 0.$$

Hence, $h(\lambda) = \lambda^{\frac{p-1}{4}}\,{}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\,1-\frac{1}{\lambda}\right)$ satisfies

$$(\lambda^3 - \lambda^2)\frac{d^2h}{d\lambda^2} + \left(\left(2 - \frac{p}{2}\right)\lambda^2 + \left(\frac{p}{2} - 1\right)\lambda\right)\frac{dh}{d\lambda} + \left(\left(\frac{p^2-4p+3}{16}\right)\lambda + -\frac{p^2}{16} + \frac{1}{36}\right)h = 0. \tag{3-3}$$

The function $h(\lambda)$ is a Laurent series in $\frac{1}{\lambda}$ with p-integral rational coefficients. However, its reduction modulo p yields a polynomial in λ. This polynomial must satisfy the reduction of (3-3) modulo p, so $F(\lambda) = \lambda^{\frac{p-1}{4}}\,{}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\,1-\frac{1}{\lambda}\right)_p$ satisfies

$$(\lambda^3 - \lambda^2)\frac{d^2F}{d\lambda^2} + (2\lambda^2 - \lambda)\frac{dF}{d\lambda} + \left(\frac{3}{16}\lambda + \frac{1}{36}\right)F \equiv 0 \pmod{p}.$$

A similar calculation shows that $F(\lambda) = \lambda^{\frac{p-3}{4}}\,{}_2F_1\left(\begin{matrix}\frac{7}{12} & \frac{11}{12}\\ & 1\end{matrix}\,\middle|\,1-\frac{1}{\lambda}\right)_p$ also satisfies the same differential equation when $p \equiv 7, 11 \pmod{12}$.

Now, to compute the Hasse invariant, we consider a general x^{p-1} term in the expansion of $(4x^3 - 27\lambda x - 27\lambda)^{\frac{p-1}{2}}$. This is of the form

$$(4x^3)^n(-27\lambda x)^{p-3n-1}(-27\lambda)^{2n-\frac{p-1}{2}},$$

where $\frac{p-1}{4} \le n \le \left\lfloor \frac{p}{3} \right\rfloor$. For a given n in this range, there are exactly

$$\binom{\frac{p-1}{2}}{n}\binom{\frac{p-1}{2}-n}{p-3n-1}$$

ways to choose which of the $4x^3-27\lambda x-27\lambda$ factors the $4x^3$ terms and $-27\lambda x$ terms came from. Summing over all n yields the Hasse invariant to be

$$\sum_{n=\frac{p-1}{4}}^{\lfloor \frac{p}{3} \rfloor} 4^n(-27\lambda)^{\frac{p-1}{2}-n}\binom{\frac{p-1}{2}}{n}\binom{\frac{p-1}{2}-n}{p-3n-1},$$

into which we can substitute $n=\frac{p-1}{2}-k$, and using the fact that $4^{\frac{p-1}{2}} \equiv 1 \pmod{p}$, we obtain

$$\sum_{k=\frac{p-1}{2}-\lfloor \frac{p}{3} \rfloor}^{\frac{p-1}{4}} \left(-\frac{27}{4}\lambda\right)^k \binom{\frac{p-1}{2}}{k}\binom{k}{3k-\frac{p-1}{2}}.$$

We show the Hasse invariant satisfies the differential equation by showing that for any t, the λ^t term in the resulting expansion is congruent to 0 mod p. Let

$$c(k)=\left(-\frac{27}{4}\lambda\right)^k \binom{\frac{p-1}{2}}{k}\binom{k}{3k-\frac{p-1}{2}}.$$

Then the λ^t term has coefficient

$$\frac{d^2}{dt^2}c(t-1)-\frac{d^2}{dt^2}c(t)+2\frac{d}{dt}c(t-1)-\frac{d}{dt}c(t)+\frac{3}{16}c(t-1)+\frac{1}{36}c(t),$$

which we expand to obtain

$$\left(-\frac{27}{4}\right)^t \binom{\frac{p-1}{2}}{t}\binom{t}{3t-\frac{p-1}{2}}\left(-t(t-1)-t+\frac{1}{36}\right)$$
$$+\left(-\frac{27}{4}\right)^{t-1} \binom{\frac{p-1}{2}}{t-1}\binom{t-1}{3t-3-\frac{p-1}{2}}\left((t-1)(t-2)+2(t-1)+\frac{3}{16}\right).$$

This is congruent to 0 modulo p if and only if

$$\binom{\frac{p-1}{2}}{t}\binom{t}{3t-\frac{p-1}{2}}\left(\frac{27}{4}t^2-\frac{3}{16}\right)+\binom{\frac{p-1}{2}}{t-1}\binom{t-1}{3t-3-\frac{p-1}{2}}\left(t^2-t+\frac{3}{16}\right)$$

is also congruent to 0. We now expand the first binomials to obtain

$$\frac{1}{t!}\left(\frac{p-1}{2}\right)\cdots\left(\frac{p-1}{2}-t+1\right)\binom{t}{3t-\frac{p-1}{2}}\left(\frac{27}{4}t^2-\frac{3}{16}\right)$$
$$+\frac{1}{(t-1!)}\left(\frac{p-1}{2}\right)\cdots\left(\frac{p-1}{2}-t+2\right)\binom{t-1}{3t-3-\frac{p-1}{2}}\left(t^2-t+\frac{3}{16}\right),$$

which is congruent to 0 modulo p if and only if

$$\frac{\frac{1}{2}-t}{t}\binom{t}{3t-\frac{p-1}{2}}\left(\frac{27}{4}t^2-\frac{3}{16}\right)+\binom{t-1}{3t-3-\frac{p-1}{2}}\left(t^2-t+\frac{3}{16}\right)\equiv 0 \pmod p$$

as well. Using a similar cancellation method on the remaining binomials shows that it is sufficient to prove

$$\left(\frac{1}{2}-t\right)\left(\frac{p-1}{2}-2t+2\right)\left(\frac{p-1}{2}-2t+1\right)\left(\frac{27}{4}t^2-\frac{3}{16}\right)$$
$$+\left(3t-\frac{p-1}{2}\right)\left(3t-\frac{p-1}{2}-1\right)\left(3t-\frac{p-1}{2}-2\right)\left(t^2-t+\frac{3}{16}\right)\equiv 0 \pmod p,$$

which is easily verified.

Thus the Hasse invariant satisfies the same second order differential equation as both $\lambda^{\frac{p-1}{4}}\ {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p$ and $\lambda^{\frac{p-3}{4}}\ {}_2F_1\left(\begin{matrix}\frac{7}{12} & \frac{11}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p$. For $p>5$, notice that both the Hasse invariant and the truncated hypergeometric functions have no term with a degree less than 2. For each case, this implies that the truncated polynomials are congruent modulo p to the Hasse invariant up to multiplication by a constant. For the case $p=5$, it is easy to compute that $\lambda\ {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_5=\lambda$, and the Hasse invariant is 4λ, so this property still holds.

Therefore, we know that the two truncated hypergeometric functions have the same roots modulo p as the Hasse invariant, so by Lemma 2.2, λ is a root of the hypergeometric functions if and only if $E_{\frac{1}{12}}(\lambda)$ is supersingular. Notice that $\lambda^{\lfloor\frac{p}{12}\rfloor}\ {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p$ and $\lambda^{\lfloor\frac{p}{12}\rfloor}\ {}_2F_1\left(\begin{matrix}\frac{7}{12} & \frac{11}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p$ have the same roots as $\lambda^{\frac{p-1}{4}}$ multiplied by the respective truncated functions with the exception of 0, which is as desired since $E_{\frac{1}{12}}(0)$ is singular. Also, when $p\equiv 1,5 \pmod{12}$ the degree of $\lambda^{\lfloor\frac{p}{12}\rfloor}\ {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p$ is $\lfloor\frac{p}{12}\rfloor$, so by Lemma 3.2, there exists a constant c_p such that

$$S_{p,\frac{1}{12}}\equiv {c_p}^{-1}\lambda^{\lfloor\frac{p}{12}\rfloor}\ {}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\middle| 1-\frac{1}{\lambda}\right)_p \pmod p.$$

Similarly for primes $p \equiv 7, 11 \pmod{12}$,

$$S_{p,\frac{1}{12}} \equiv {c_p}^{-1} \lambda^{\lfloor \frac{p}{12} \rfloor} \, {}_2F_1\left(\begin{matrix} \frac{7}{12} & \frac{11}{12} \\ & 1 \end{matrix} \,\middle|\, 1-\frac{1}{\lambda}\right)_p \pmod{p}.$$

Finally, we explicitly compute the constant c_p. Notice that $S_{p,\frac{1}{12}}$ is monic, so c_p is the coefficient of the leading term in $\lambda^{\lfloor \frac{p}{12} \rfloor} \, {}_2F_1\left(\begin{matrix} \frac{1}{12} & \frac{5}{12} \\ & 1 \end{matrix} \,\middle|\, 1-\frac{1}{\lambda}\right)_p$, the same as the constant term in ${}_2F_1\left(\begin{matrix} \frac{1}{12} & \frac{5}{12} \\ & 1 \end{matrix} \,\middle|\, 1-\frac{1}{\lambda}\right)_p$. For $n > \lfloor \frac{p}{12} \rfloor$, one of $\left(\frac{1}{12}\right)_n$ or $\left(\frac{5}{12}\right)_n$ will be congruent to 0 modulo p. Hence, the constant term of

$${}_2F_1\left(\begin{matrix} \frac{1}{12} & \frac{5}{12} \\ & 1 \end{matrix} \,\middle|\, 1-\frac{1}{\lambda}\right)_p = \sum_{n=0}^{\lfloor \frac{p}{12} \rfloor} \frac{\left(\frac{1}{12}\right)_n \left(\frac{5}{12}\right)_n}{n!^2} \left(1-\frac{1}{\lambda}\right)^n$$

is

$$\sum_{n=0}^{\lfloor \frac{p}{12} \rfloor} \frac{\left(\frac{1}{12}\right)_n \left(\frac{5}{12}\right)_n}{n!^2}.$$

For $p \equiv 1 \pmod{12}$, we have

$$\frac{\left(\frac{1}{12}\right)_n}{n!} \equiv (-1)^n \frac{\frac{p-1}{12}\frac{p-13}{12}\cdots\left(\frac{p-1}{12}-n+1\right)}{n!} \pmod{p} \equiv (-1)^n \binom{\frac{p-1}{12}}{n} \pmod{p}.$$

Also, $\frac{\left(\frac{5}{12}\right)_n}{n!} \equiv (-1)^n \binom{\frac{5p-5}{12}}{n} \pmod{p}$. Therefore,

$$c_p = \sum_{n=0}^{\lfloor \frac{p}{12} \rfloor} \frac{\left(\frac{1}{12}\right)_n \left(\frac{5}{12}\right)_n}{n!^2} \equiv \binom{6\lfloor \frac{p}{12} \rfloor}{\lfloor \frac{p}{12} \rfloor} \pmod{p}.$$

For $p \equiv 5 \pmod{12}$,

$$c_p = \sum_{n=0}^{\lfloor \frac{p}{12} \rfloor} \frac{\left(\frac{1}{12}\right)_n \left(\frac{5}{12}\right)_n}{n!^2} \equiv \binom{6\lfloor \frac{p}{12} \rfloor + 2}{\lfloor \frac{p}{12} \rfloor} \pmod{p}.$$

A similar method can be used to compute $c_p \equiv \binom{6\lfloor \frac{p}{12} \rfloor + 2}{\lfloor \frac{p}{12} \rfloor} \pmod{p}$ when $p \equiv 7 \pmod{12}$ and $\binom{6\lfloor \frac{p}{12} \rfloor + 4}{\lfloor \frac{p}{12} \rfloor}$ when $p \equiv 11 \pmod{12}$, which completes the proof. □

Proof of Corollary 1.3. Recall from the proof of Theorem 1.2 that since the Hasse invariant of $E_{\frac{1}{12}}(\lambda)$ and the polynomial $\lambda^{\frac{p-1}{4}} \, {}_2F_1\left(\begin{matrix} \frac{1}{12} & \frac{5}{12} \\ & 1 \end{matrix} \,\middle|\, 1-\frac{1}{\lambda}\right)_p$ both satisfied the same second order differential equation, they are congruent up to multiplication by a constant, which we will denote b_p. The same argument and notation apply to $\lambda^{\frac{p-3}{4}}\,{}_2F_1\left(\begin{matrix}\frac{7}{12} & \frac{11}{12}\\ & 1\end{matrix}\,\middle|\, 1-\frac{1}{\lambda}\right)_p$ when $p \equiv 7, 11 \mod 12$).

Assume that $p \equiv 1 \mod 12$, and define $m = \left\lfloor \frac{p}{12} \right\rfloor$. Also, define $n = 3m - k$. We computed the Hasse invariant of $E_{\frac{1}{12}}(\lambda)$ to be

$$\sum_{k=2m}^{3m} \left(\frac{-27}{4}\lambda\right)^k \binom{6m}{k}\binom{k}{3m-6m} = \sum_{n=0}^{m} \left(\frac{-27\lambda}{4}\right)^{3m-n} \binom{6m}{3m-n}\binom{3m-n}{3m-3n}.$$

By definition,

$$\lambda^{\frac{p-1}{4}}\,{}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\, 1-\frac{1}{\lambda}\right)_p \equiv \lambda^{\frac{p-1}{4}} \sum_{k=0}^{m} \frac{\left(\frac{1}{12}\right)_k \left(\frac{5}{12}\right)_k}{k!^2}\left(1-\frac{1}{\lambda}\right)^k \pmod{p}.$$

As before,

$$\frac{\left(\frac{1}{12}\right)_k \left(\frac{5}{12}\right)_k}{k!^2} \equiv \binom{m}{k}\binom{5m}{k} \pmod{p}.$$

We expand each of the $\left(1-\frac{1}{\lambda}\right)^k$ terms and rearrange to obtain

$$\begin{aligned}\lambda^{\frac{p-1}{4}}\,{}_2F_1\left(\begin{matrix}\frac{1}{12} & \frac{5}{12}\\ & 1\end{matrix}\,\middle|\, 1-\frac{1}{\lambda}\right)_p &\equiv \sum_{k=2m}^{3m} (-\lambda)^k \sum_{t=3m-k}^{m} \binom{m}{t}\binom{5m}{t}\binom{t}{3m-k} \pmod{p}\\ &\equiv \sum_{n=0}^{m} (-\lambda)^{3m-n} \sum_{t=n}^{m} \binom{m}{t}\binom{5m}{t}\binom{t}{n} \pmod{p}.\end{aligned}$$

Since this polynomial is congruent to the Hasse invariant via multiplication by b_p, we have, for all $0 \le n \le m$,

$$\left(\frac{27}{4}\right)^{3m-n} \binom{3m-n}{3m-3n}\binom{6m}{3m-n} \equiv b_p \sum_{t=n}^{m} \binom{m}{t}\binom{5m}{t}\binom{t}{n} \pmod{p}.$$

When $n = 0$, this becomes

$$\left(\frac{27}{4}\right)^{3m} \binom{6m}{3m} \equiv b_p \sum_{t=0}^{m} \binom{m}{t}\binom{5m}{t} \equiv b_p \binom{6m}{m} \pmod{p}$$

and thus

$$b_p \equiv \frac{\binom{6m}{3m}\left(\frac{27}{4}\right)^{3m}}{\binom{6m}{m}} \pmod{p}.$$

Substituting this back into our identity, we have that for all $0 \le n \le m$,

$$\left(\frac{4}{27}\right)^n \binom{3m-n}{3m-3n}\binom{6m}{3m-n}\binom{6m}{m} \equiv \binom{6m}{3m}\sum_{t=n}^{m}\binom{m}{t}\binom{5m}{t}\binom{t}{n} \pmod{p}.$$

In the case $n = m$, we obtain the simpler identity

$$\left(\frac{27}{4}\right)^{3m}\binom{5m}{m}\binom{6m}{3m} \equiv \binom{6m}{2m}\binom{6m}{m} \pmod{p}. \qquad \square$$

4. Examples

In this section we provide two examples to illustrate our main theorems.

Example of Theorem 1.1. Consider $p = 19$. The supersingular j-invariants mod 19 are known to be 18 (corresponding to 1728) and 7. From formula (3-1) we find that the values of λ where $j \equiv 18 \pmod{19}$ are $-1 \pm i\sqrt{6}$ only. The values of λ for which $j \equiv 7 \pmod{19}$ are $-6 \pm 3\sqrt{2}$ and $4 \pm 11\sqrt{13}$. Thus

$$\begin{aligned} S_{19,\frac{1}{3}}(\lambda) &= (\lambda-(-1+i\sqrt{6}))(\lambda-(-1-i\sqrt{6}))(\lambda-(-6+3\sqrt{2})) \\ &\quad (\lambda-(-6-3\sqrt{2}))(\lambda-(4+11\sqrt{13}))(\lambda-(4-11\sqrt{13})) \\ &\equiv \lambda^6+6\lambda^5+14\lambda^4+8\lambda^3+13\lambda^2+5\lambda+12 \pmod{19} \\ &\equiv (\lambda^2+2\lambda+7)(\lambda^2+11\lambda+1)(\lambda^2+12\lambda+18) \pmod{19}. \end{aligned}$$

The Hasse invariant is the coefficient of $(wxy)^{18}$ in the expansion of

$$(wy^2+\lambda wxy+\lambda^2w^2y-x^3)^{18}.$$

This is

$$H(\lambda) \equiv \lambda^{18}+6\lambda^{17}+14\lambda^{16}+8\lambda^{15}+13\lambda^{14}+5\lambda^{13}+12\lambda^{12} \equiv \lambda^{12}\, S_{19,\frac{1}{3}}(\lambda) \pmod{19}.$$

In addition,

$${}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3} \\ & 1\end{matrix} \,\middle|\, \frac{27}{\lambda}\right)_{19} \equiv 1+\frac{6}{\lambda}+\frac{14}{\lambda^2}+\frac{8}{\lambda^3}+\frac{13}{\lambda^4}+\frac{5}{\lambda^5}+\frac{12}{\lambda^6} \equiv \frac{1}{\lambda^6}S_{19,\frac{1}{3}}(\lambda) \pmod{19}.$$

Example of Theorem 1.2. Consider $p = 59$, which is 11 modulo 12. The supersingular j-invariants mod 59 are known to be 0, 17 (corresponding to 1728), 48, 47, 28, and 15. From formula (3-2), we find the λ-invariants corresponding to 48, 47, 28, and 15 are 32, 35, 24, and 22, respectively. We do not include the cases $j = 0$ or $j = 1728$ since in these cases $E_{\frac{1}{12}}(\lambda)$ is singular. Thus

$$\begin{aligned} S_{59,\frac{1}{12}}(\lambda) &= (\lambda+27)(\lambda+24)(\lambda+35)(\lambda+37) \\ &\equiv \lambda^4+5\lambda^3+10\lambda^2+11\lambda+3 \pmod{59}. \end{aligned}$$

The Hasse invariant is the coefficient of x^{58} in $(4x^3-27\lambda x-27\lambda)^{29}$. This is

$$H(\lambda)\equiv 2\lambda^{14}+10\lambda^{13}+20\lambda^{12}+22\lambda^{11}+6\lambda^{10}\equiv 2\lambda^{10}S_{59,\frac{1}{12}}(\lambda) \pmod{59}.$$

In addition,

$${}_2F_1\left(\begin{matrix}\frac{7}{12} & \frac{11}{12}\\ & 1\end{matrix}\,\middle|\, 1-\frac{1}{\lambda}\right)_{59}\equiv 2+\frac{10}{\lambda}+\frac{20}{\lambda^2}+\frac{22}{\lambda^3}+\frac{6}{\lambda^4}\equiv\frac{2}{\lambda^4}S_{59,\frac{1}{12}}(\lambda) \pmod{59} \pmod{59}.$$

Also, $c_{59}\equiv\binom{28}{4}\equiv 2 \pmod{59}$.

5. Conclusion

We have described the supersingular loci of two infinite families of elliptic curves in terms of truncated hypergeometric functions. For the family $E_{\frac{1}{3}}(\lambda)$, the supersingular locus was a power of λ times the ${}_2F_1\left(\begin{matrix}\frac{1}{3} & \frac{2}{3}\\ & 1\end{matrix}\,\middle|\,\frac{27}{\lambda}\right)_p$ function. We found a similar result for the family $E_{\frac{1}{12}}(\lambda)$. This gives a very simple method for determining exactly which values of λ yield supersingular curves for these infinite families. Over any given field $\mathbb{F}_p$, these λ-invariants are simply the roots of these hypergeometric functions truncated modulo p.

Our results also yield interesting insights into combinatorics. We have the very nice identity given in Corollary 1.3, and analogous results can be obtained by similar methods. For example, assume that p is any prime that is congruent to 1 modulo 12 and that $12m+1=p$. If one could prove that the constant b_p from the proof of Corollary 1.3 is congruent to 1 modulo p for all such p, then the following identity is implied from Corollary 1.3:

$$\binom{6m}{3m}\equiv\left(\frac{27}{4}\right)^m\binom{2m}{m} \pmod{p}.$$

The truth of this statement has been verified for all m up to 10000. This is a fascinating identity regarding the "central" binomial coefficients modulo p, and it illustrates the types of insights one can gain into combinatorics through the study of elliptic curves and hypergeometric functions.

It is our hope that these results will be used to further understand the deep connections between elliptic curves and hypergeometric functions.

References

[Borwein and Borwein 1987] J. M. Borwein and P. B. Borwein, *Pi and the AGM: a study in analytic number theory and computational complexity*, Wiley, New York, 1987. MR 89a:11134 Zbl 0611.10001

[El-Guindy and Ono 2012] A. El-Guindy and K. Ono, “Hasse invariants for the Clausen elliptic curves”, preprint, 2012, available at http://www.mathcs.emory.edu/~ono/publications-cv/pdfs/129.pdf. To appear in *Ramanujan J.*

[Husemöller 2004] D. Husemöller, *Elliptic curves*, 2nd ed., Graduate Texts in Mathematics **111**, Springer, New York, 2004. MR 2005a:11078 Zbl 1040.11043

[Lennon 2010] C. Lennon, “A trace formula for certain Hecke operators and Gaussian hypergeometric functions”, preprint, 2010. arXiv 1003.1157

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Washington 2003] L. C. Washington, *Elliptic curves: number theory and cryptography*, Chapman & Hall/CRC, Boca Raton, FL, 2003. MR 2004e:11061 Zbl 1034.11037

Received: 2011-09-12 Accepted: 2011-09-14

monks@college.harvard.edu *Harvard University, 2013 Harvard Yard Mail Center,*
Cambridge 02138, 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

2012 vol. 5 no. 1

1944-4176(2012)5:1;1-C