involve

a journal of mathematics

mathematical sciences publishers

2009 Vol. 2, No. 1

involve
pjm.math.berkeley.edu/involve

See inside back cover or http://pjm.math.berkeley.edu/involve for submission instructions and subscription prices.
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, at Mathematical Sciences Publisher, 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.

PUBLISHED BY
mathematical sciences publishers
http://www.mathscipub.org
A NON-PROFIT CORPORATION
Typeset in LaTeX

INVOLVE 2:1(2009)

Oscillation criteria for two-dimensional systems of first-order linear dynamic equations on time scales

Douglas R. Anderson and William R. Hall

(Communicated by John V. Baxley)

Oscillation criteria for two-dimensional difference systems of first-order linear difference equations are generalized and extended to arbitrary dynamic equations on time scales. This unifies under one theory corresponding results from differential systems, and includes second-order self-adjoint differential, difference, and q-difference equations within its scope. Examples are given illustrating a key theorem.

1. Prelude

Jiang and Tang [2007] have established sufficient conditions for the oscillation of the linear two-dimensional difference system

$$\Delta x_n = p_n y_n, \quad \Delta y_{n-1} = -q_n x_n, \quad n \in \mathbb{Z}, \tag{1-1}$$

where $\{p_n\}$, $\{q_n\}$ are nonnegative real sequences and Δ is the forward difference operator given via $\Delta x_n = x_{n+1} - x_n$; see also [Li 2001]. The system (1-1) may be viewed as a discrete analogue of the differential system

$$x'(t) = p(t)y(t), \quad y'(t) = -q(t)x(t), \quad t \in \mathbb{R}, \tag{1-2}$$

investigated in [Lomtatidze and Partsvania 1999].

Oscillation questions in difference and differential equations are an interesting and important area of study in modern mathematics. Furthermore, within the past two decades, these two related but distinct areas have begun to be combined under a powerful, more robust and general theory titled dynamic equations on time scales, a theory introduced by Hilger [1990]. For example, equations (1-1) and (1-2) would take the form

$$x^{\Delta}(t) = p(t)y(t), \quad y^{\nabla}(t) = -q(t)x(t), \quad t \in \mathbb{T}, \tag{1-3}$$

MSC2000: primary 34B10; secondary 39A10.
Keywords: oscillation, linear system, time scales.

where $\mathbb{T}$ is an arbitrary time scale (any nonempty closed set of real numbers) unbounded above, with the special cases of $\mathbb{T}=\mathbb{Z}$ and $\mathbb{T}=\mathbb{R}$ yielding systems (1-1) and (1-2), respectively, as important corollaries. In this general time-scale setting, Δ represents the delta (or Hilger) derivative [Bohner and Peterson 2001, Definition 1.10], and ∇ represents the nabla derivative, introduced in [Atici and Guseinov 2002, Section 2]:

$$x^{\Delta}(t) := \lim_{s\to t} \frac{x(\sigma(t))-x(s)}{\sigma(t)-s}, \qquad x^{\nabla}(t) := \lim_{s\to t} \frac{x(\rho(t))-x(s)}{\rho(t)-s},$$

where $\sigma(t) := \inf\{s\in\mathbb{T} : s>t\}$ is the forward jump operator and $\rho(t) := \sup\{s\in\mathbb{T} : s<t\}$ is the backward jump operator. Moreover, $\mu(t) := \sigma(t)-t$ is the forward graininess function, and $\nu(t) := t-\rho(t)$ is the backward graininess function. In particular, if $\mathbb{T}=\mathbb{R}$, then $\sigma(t)=t=\rho(t)$ and $x^{\Delta}=x'=x^{\nabla}$, while if $\mathbb{T}=h\mathbb{Z}$ for any $h>0$, then $\sigma(t)=t+h$ and $\rho(t)=t-h$, so that

$$x^{\Delta}(t) = \frac{x(t+h)-x(t)}{h} \quad \text{and} \quad x^{\nabla}(t) = \frac{x(t)-x(t-h)}{h},$$

respectively. A function $f:\mathbb{T}\to\mathbb{R}$ is right-dense continuous provided it is continuous at each right-dense point $t\in\mathbb{T}$ (a point where $\sigma(t)=t$) and has a left-sided limit at each left-dense point $t\in\mathbb{T}$ (a point where $\rho(t)=t$). The set of right-dense continuous functions on $\mathbb{T}$ is denoted by $\mathrm{C_{rd}}(\mathbb{T})$. It can be shown that any right-dense continuous function f has an antiderivative (a function $F:\mathbb{T}\to\mathbb{R}$ with the property $F^{\Delta}(t)=f(t)$ for all $t\in\mathbb{T}$). The Cauchy delta integral of f is defined by

$$\int_{t_0}^{t_1} f(t)\Delta t = F(t_1)-F(t_0),$$

where F is an antiderivative of f on $\mathbb{T}$. Similar notions hold for left-dense continuous functions and the Cauchy nabla integral. For example, if $\mathbb{T}=\mathbb{Z}$, then

$$\int_{t_0}^{t_1} f(t)\Delta t = \sum_{t=t_0}^{t_1-1} f(t) \quad \text{and} \quad \int_{t_0}^{t_1} f(t)\nabla t = \sum_{t=t_0+1}^{t_1} f(t),$$

and if $\mathbb{T}=\mathbb{R}$, then

$$\int_{t_0}^{t_1} f(t)\Delta t = \int_{t_0}^{t_1} f(t)dt = \int_{t_0}^{t_1} f(t)\nabla t.$$

Throughout we assume that $t_0<t_1$ are points in $\mathbb{T}$, and define the time-scale interval $[t_0,t_1]_{\mathbb{T}} = \{t\in\mathbb{T} : t_0\le t\le t_1\}$. Other time-scale intervals are defined similarly. For convenience, the composition $x\circ\sigma$ is denoted x^{σ}, and $x\circ\rho$ is denoted x^{ρ}. For more on time scales and time-scale notation, see the fundamental texts [Bohner and Peterson 2001; 2003].

System (1-3) is a generalization of a key second-order linear dynamic equation. To see this, suppose the potential p is nabla differentiable and strictly positive. Then we have

$$x^{\Delta\nabla}(t) = [p(t)y(t)]^{\nabla} = p^{\rho}(t)y^{\nabla}(t) + p^{\nabla}(t)y(t) = -p^{\rho}(t)q(t)x(t) + \frac{p^{\nabla}(t)}{p(t)}x^{\Delta}(t),$$

which we can rewrite in the (formally) self-adjoint form

$$\Big(\frac{1}{p}x^{\Delta}\Big)^{\nabla}(t) + q(t)x(t) = 0, \quad t \in \mathbb{T}; \tag{1-4}$$

see [Bohner and Peterson 2003, Section 4.3]. Thus the system (1-3) is an extension of the second-order self-adjoint Equation (1-4), and many important equations are included under the rubric of our discussion below, including the second-order self-adjoint differential equation

$$\Big(\frac{1}{p}x'\Big)'(t) + q(t)x(t) = 0, \quad t \in \mathbb{R},$$

the second-order self-adjoint difference equation

$$\Delta\Big(\frac{1}{p_{n-1}}\Delta x_{n-1}\Big) + q_n x_n = 0, \quad n \in \mathbb{Z},$$

and the second-order self-adjoint q-difference (quantum) equation ($q > 1$)

$$D^q\Big(\frac{1}{p(t)}D_q x(t)\Big) + q(t)x(t) = 0, \quad t \in q^{\mathbb{Z}},$$

where

$$D^q x(t) = \frac{x(t) - x(t/q)}{t - t/q} \quad \text{and} \quad D_q x(t) = \frac{x(qt) - x(t)}{qt - t} \tag{1-5}$$

are the quantum backward and forward derivatives, respectively.

2. Preliminary results on oscillation

Let $\mathbb{T}$ be a time scale that is unbounded above, and let $t_0 \in \mathbb{T}$. In (1-3), assume $p : \mathbb{T} \to \mathbb{R}$ is right-dense continuous with $p > 0$ on $[t_0, \infty)_{\mathbb{T}}$, and $q : \mathbb{T} \to \mathbb{R}$ is continuous with $q \geq 0$ on $[t_0, \infty)_{\mathbb{T}}$; then p is delta integrable and q is integrable. Note the stronger continuity condition on the potential q; from the right-hand equation in (1-3), we then have that y^{∇} is continuous, so that y is delta differentiable as well, with $y^{\Delta} = y^{\nabla\sigma} = -q^{\sigma}x^{\sigma}$. An alternative approach would be to use only delta derivatives in (1-3), with p and q both right-dense continuous functions. The results in the sequel would be analogous to those derived below, but would not incorporate the self-adjoint form (1-4), nor directly extend (1-1). Our techniques are modelled after those found in [Jiang and Tang 2007; Lomtatidze and Partsvania 1999] and the references therein.

A solution (x, y) of (1-3) is oscillatory if both component functions x and y are oscillatory, that is to say neither eventually positive nor eventually negative; otherwise, the solution is nonoscillatory. The dynamic system (1-3) is oscillatory if all its solutions are oscillatory.

Lemma 2.1. *The component functions x and y of a nonoscillatory solution (x, y) of (1-3) are themselves nonoscillatory.*

Proof. Assume to the contrary that x oscillates but y is eventually positive. Then $x^\Delta = py > 0$ eventually, so that $x(t) > 0$ or $x(t) < 0$ for all large $t \in \mathbb{T}$, a contradiction. The case where y is eventually negative is similar. Likewise, assuming that y oscillates while x is eventually positive or eventually negative leads to comparable contradictions. □

Lemma 2.2. *If*

$$\int_{t_0}^{\infty} p(r)\Delta r = \infty \quad \textit{and} \quad \int_{\rho(t_0)}^{\infty} q(s)\nabla s = \infty, \tag{2-1}$$

then each solution of (1-3) is oscillatory.

Proof. Let (x, y) be a nonoscillatory solution of (1-3). Without loss of generality, we may assume that $x > 0$; then $y^\nabla = -qx \le 0$, and in view of Lemma 2.1, y must be of constant sign eventually. If $y(t_1) < 0$ for some $t_1 \in [t_0, \infty)_\mathbb{T}$, then $y < 0$ on $[t_1, \infty)_\mathbb{T}$ and $x^\Delta = py < 0$ on $[t_1, \infty)_\mathbb{T}$; after delta integrating from t_1 to t, we have

$$x(t) = x(t_1) + \int_{t_1}^{t} p(r)y(r)\Delta r.$$

Since y is negative and nonincreasing, by the first assumption in (2-1) the right-hand side tends to $-\infty$, in contradiction with $x > 0$. Consequently, $y > 0$ with $y^\nabla \le 0$ on $[t_0, \infty)_\mathbb{T}$, and $x^\Delta > 0$ on $[t_0, \infty)_\mathbb{T}$ by the first equation of (1-3). Thus there exists a constant $c > 0$ and $t_1 \in [t_0, \infty)_\mathbb{T}$ such that $x(t) \ge c$ for $t \in [t_1, \infty)_\mathbb{T}$. Nabla integrating the second equation of (1-3), we obtain

$$c \int_{t_1}^{\infty} q(s)\nabla s \le y(t_1) < \infty,$$

and this contradicts the second assumption in (2-1). □

Lemma 2.3. *If*

$$\int_{t_0}^{\infty} p(r)\Delta r < \infty \quad \textit{and} \quad \int_{\rho(t_0)}^{\infty} q(s)\nabla s < \infty, \tag{2-2}$$

then the system (1-3) is nonoscillatory.

Proof. Suppose that (2-2) holds. Then there exists $t_1 \in [t_0, \infty)_{\mathbb{T}}$ such that

$$\int_{t_1}^{\infty} p(r)\left(1+2\int_{r}^{\infty} q(s)\nabla s\right)\Delta r < 1. \tag{2-3}$$

Let $\mathscr{B} = \mathrm{C}_{\mathrm{rd}}(\mathbb{T})$ be the Banach space of right-dense continuous functions on $\mathbb{T}$, with norm $\|x\| = \sup_{t \geq t_1, t\in\mathbb{T}} |x(t)|$ and the usual pointwise ordering $\leq$. Define a subset $\mathscr{S}$ of $\mathscr{B}$ by

$$\mathscr{S} = \left\{x \in \mathscr{B} : 1 \leq x(t) \leq 2,\ t \in [t_1, \infty)_{\mathbb{T}}\right\}.$$

For any subset $\mathscr{Q}$ of $\mathscr{S}$, we have that $\inf \mathscr{Q} \in \mathscr{S}$ and $\sup \mathscr{Q} \in \mathscr{S}$. Let $L : \mathscr{S} \to \mathscr{B}$ be the functional given via

$$(Lx)(t) = 1 + \int_{t_1}^{t} p(r)\left(1+\int_{r}^{\infty} q(s)x(s)\nabla s\right)\Delta r, \quad t \in [t_1, \infty)_{\mathbb{T}}.$$

By the assumptions on $x \in \mathscr{S}$ and p and q, $(Lx)(t) \geq 1$ for all $t \in [t_1, \infty)_{\mathbb{T}}$, and

$$(Lx)(t) \leq 1 + \int_{t_1}^{t} p(r)\left(1+\int_{r}^{\infty} 2q(s)\nabla s\right)\Delta r \leq 2$$

by (2-3). Moreover,

$$(Lx)^{\Delta}(t) = p(t)\left(1+\int_{t}^{\infty} q(s)x(s)\nabla s\right) > 0, \tag{2-4}$$

ensuring that $L : \mathscr{S} \to \mathscr{S}$ is increasing. By Knaster's fixed-point theorem [Knaster 1928], we can conclude that there exists an $x \in \mathscr{S}$ such that $x = Lx$. If we let

$$y(t) = 1 + \int_{t}^{\infty} q(s)x(s)\nabla s, \quad t \in [t_1, \infty)_{\mathbb{T}}$$

using the fixed point $x \in \mathscr{S}$, then we have

$$x^{\Delta}(t) = (Lx)^{\Delta}(t) = p(t)y(t) \quad \text{and} \quad y^{\nabla}(t) = -q(t)x(t)$$

for $t \in [t_1, \infty)_{\mathbb{T}}$ by using (2-4). Thus (x, y) is a nonoscillatory solution of (1-3). □

In light of Lemmas 2.2 and 2.3, respectively, we could assume that either

$$\int_{t_0}^{\infty} p(r)\Delta r = \infty \quad \text{and} \quad \int_{\rho(t_0)}^{\infty} q(s)\nabla s < \infty \tag{2-5}$$

or

$$\int_{t_0}^{\infty} p(r)\Delta r < \infty \quad \text{and} \quad \int_{\rho(t_0)}^{\infty} q(s)\nabla s = \infty;$$

in fact, we will focus on (2-5). Moreover, in preparation for what follows, we introduce the following notation. Let

$$P(t) := \int_{t_0}^{t} p(r)\Delta r. \tag{2-6}$$

Lemma 2.4. *Assume that (2-5) holds, P is given by (2-6), and $\lambda \in [0, 1)$ is a real number. If*

$$\lim_{t\to\infty} \frac{\mu(t)p(t)}{P(t)} = 0, \quad \left(\textit{equivalently}, \quad \lim_{t\to\infty} \frac{P^\sigma(t)}{P(t)} = 1\right) \tag{2-7}$$

then given $\varepsilon > 0$ there exists a $t_1 \equiv t_1(\varepsilon) \in (t_0, \infty)_{\mathbb{T}}$ such that for any $t \in [t_1, \infty)_{\mathbb{T}}$,

$$\int_t^\infty \frac{\left[(P^\lambda)^\Delta(r)\right]^2}{p(r)P^\lambda(r)}\Delta r \le \frac{\lambda^2}{1-\lambda}(1+\varepsilon)^{2-\lambda} P^{\lambda-1}(t), \quad \textit{and} \tag{2-8}$$

$$\int_t^\infty \frac{p(r)}{P^{2-\lambda}(r)}\Delta r \le \frac{(1+\varepsilon)^{2-\lambda}}{1-\lambda} P^{\lambda-1}(t). \tag{2-9}$$

Proof. For $r \in (t_0, \infty)_{\mathbb{T}}$, by the chain rule [Bohner and Peterson 2001, Theorem 1.90] we have

$$(P^\lambda)^\Delta(r) = \begin{cases} \dfrac{P^\lambda(\sigma(r)) - P^\lambda(r)}{\mu(r)} & \text{if } \mu(r) > 0, \\ \lambda p(r) P^{\lambda-1}(r) & \text{if } \mu(r) = 0. \end{cases}$$

By [Bohner and Peterson 2001, Theorem 1.16 (iv)], $\mu P^\Delta = P^\sigma - P$, so that $\mu p = P^\sigma - P$ on $\mathbb{T}$. If $r \in (t_0, \infty)_{\mathbb{T}}$ is a right-scattered point, then $\mu(r) > 0$ and, suppressing the r,

$$\begin{aligned} \frac{\left[(P^\lambda)^\Delta\right]^2}{pP^\lambda} &= \frac{p}{\mu^2 p^2 P^\lambda}\left((P^\sigma)^\lambda - P^\lambda\right)^2 = \frac{p}{P^\lambda}\left(\frac{(P^\sigma)^\lambda - P^\lambda}{P^\sigma - P}\right)^2 \\ &\overset{\text{MVT}}{=} \frac{p}{P^\lambda}\left(\lambda \xi^{\lambda-1}\right)^2, \quad \xi \in \left(P(r), P^\sigma(r)\right)_{\mathbb{R}} \\ &\le \frac{p\lambda^2}{P^\lambda} P^{2\lambda-2}, \quad \lambda - 1 < 0 \\ &= \lambda^2 p P^{\lambda-2}. \end{aligned}$$

If $r \in (t_0, \infty)_{\mathbb{T}}$ is a right-dense point, then $\mu(r) = 0$ and

$$\frac{\left[(P^\lambda)^\Delta\right]^2}{pP^\lambda} = \frac{\left[\lambda p P^{\lambda-1}\right]^2}{pP^\lambda} = \lambda^2 p P^{\lambda-2}.$$

It follows that in either case,

$$\frac{\left[(P^\lambda)^\Delta(r)\right]^2}{p(r)P^\lambda(r)} \le \lambda^2 p(r)P^{\lambda-2}(r), \quad r \in (t_0,\infty)_{\mathbb{T}}. \tag{2-10}$$

Similarly, if $r \in (t_0,\infty)_{\mathbb{T}}$ is a right-scattered point, then once again $\mu(r) > 0$ and, suppressing the r,

$$\begin{aligned}-(P^{\lambda-1})^\Delta &= \frac{-p}{\mu p}\left((P^\sigma)^{\lambda-1} - P^{\lambda-1}\right) = -p\left(\frac{(P^\sigma)^{\lambda-1} - P^{\lambda-1}}{P^\sigma - P}\right)\\ &\overset{\text{MVT}}{=} p(1-\lambda)\eta^{\lambda-2}, \quad \eta \in \left(P(r), P^\sigma(r)\right)_{\mathbb{R}}\\ &\ge p(1-\lambda)(P^\sigma)^{\lambda-2}.\end{aligned}$$

If r is a right-dense point, then $P^\sigma = P$, $\mu(r)=0$, and $p(1-\lambda)P^{\lambda-2} = -(P^{\lambda-1})^\Delta$. Summarizing, in either case we have

$$-(P^{\lambda-1})^\Delta \ge p(1-\lambda)(P^\sigma)^{\lambda-2}, \quad r \in (t_0,\infty)_{\mathbb{T}}. \tag{2-11}$$

Combining (2-10) and (2-11), we see that

$$\frac{\left[(P^\lambda)^\Delta(r)\right]^2}{p(r)P^\lambda(r)} \le \frac{\lambda^2}{1-\lambda}\left(\frac{P(r)}{P^\sigma(r)}\right)^{\lambda-2}\left[-(P^{\lambda-1})^\Delta(r)\right].$$

By (2-7), given $\varepsilon > 0$ there exists a $t_1 \in [t_0,\infty)_{\mathbb{T}}$ such that $P^\sigma/P \le (1+\varepsilon)$ on $[t_1,\infty)_{\mathbb{T}}$. Consequently, for any $t \in [t_1,\infty)_{\mathbb{T}}$,

$$\begin{aligned}\int_t^\infty \frac{\left[(P^\lambda)^\Delta(r)\right]^2}{p(r)P^\lambda(r)}\Delta r \quad &\le \quad \frac{\lambda^2}{1-\lambda}(1+\varepsilon)^{2-\lambda}\int_t^\infty \left[-(P^{\lambda-1})^\Delta(r)\right]\Delta r\\ &\overset{(2\text{-}5),(2\text{-}6)}{=} \frac{\lambda^2}{1-\lambda}(1+\varepsilon)^{2-\lambda}\, P^{\lambda-1}(t),\end{aligned}$$

which is (2-8). Moreover, again for any $r \in [t_1,\infty)_{\mathbb{T}}$,

$$\begin{aligned}\frac{p(r)}{P^{2-\lambda}(r)} \quad &= \quad \frac{p(r)}{P^{2-\lambda}(\sigma(r))}\frac{P^{2-\lambda}(\sigma(r))}{P^{2-\lambda}(r)} \le (1+\varepsilon)^{2-\lambda}\frac{p(r)}{P^{2-\lambda}(\sigma(r))}\\ &\overset{(2\text{-}11)}{\le} \frac{(1+\varepsilon)^{2-\lambda}}{\lambda-1}(P^{\lambda-1})^\Delta(r).\end{aligned} \tag{2-12}$$

Delta integrating (2-12) from t to infinity, we obtain

$$\int_t^\infty \frac{p(r)}{P^{2-\lambda}(r)}\Delta r \le \frac{(1+\varepsilon)^{2-\lambda}}{\lambda-1}\int_t^\infty (P^{\lambda-1})^\Delta(r)\Delta r \overset{(2\text{-}5),(2\text{-}6)}{=} \frac{(1+\varepsilon)^{2-\lambda}}{1-\lambda}P^{\lambda-1}(t),$$

which is (2-9). □

Note that if $\mathbb{T} = \mathbb{R}$, then (2-7) is automatically satisfied, as $\mu(t) \equiv 0$.

Lemma 2.5. *Assume that (2-5) holds, that P is given by (2-6), and that (2-7) holds. If for some real number $\lambda < 1$ we have*

$$\int_{t_1}^{\infty} q^{\sigma}(r) P^{\lambda}(r) \Delta r = \infty \quad \textit{for} \quad t_1 > \sigma(t_0), \tag{2-13}$$

then the system (1-3) is oscillatory.

Proof. By Lemma 2.3, we can focus on $\lambda \in (0, 1)$. Assume that (x, y) is a nonoscillatory solution of the system (1-3); without loss of generality, assume that $x > 0$ on $[t_0, \infty)_{\mathbb{T}}$. As in the proof of Lemma 2.2, $y > 0$ with $y^{\nabla} \le 0$ and $x^{\Delta} > 0$ on $[t_0, \infty)_{\mathbb{T}}$. Let $w := y/x$. Then $w > 0$, and suppressing the argument, we have by the delta quotient rule and (1-3) that on $[t_0, \infty)_{\mathbb{T}}$,

$$w^{\Delta} = \frac{x y^{\Delta} - y x^{\Delta}}{x x^{\sigma}} = -q^{\sigma} - p w y / x^{\sigma} \le -q^{\sigma} - p w w^{\sigma} < 0. \tag{2-14}$$

In fact this gives us

$$w^{\Delta} \le -q^{\sigma} - p (w^{\sigma})^2, \tag{2-15}$$

and from the previous line we obtain on $[t_0, \infty)_{\mathbb{T}}$ that

$$\left(\frac{1}{w}\right)^{\Delta} = \frac{-w^{\Delta}}{w w^{\sigma}} \ge \frac{q^{\sigma} + p w w^{\sigma}}{w w^{\sigma}} \ge p;$$

delta integrating from t_0 to t we see that

$$1 > 1 - \frac{w(t)}{w(t_0)} \ge w(t) \int_{t_0}^{t} p(r) \Delta r = w(t) P(t) \ge 0, \quad t \in [t_0, \infty)_{\mathbb{T}}. \tag{2-16}$$

Again by the mean value theorem, $(P^{\lambda})^{\Delta} \le \lambda p P^{\lambda - 1}$ for $\lambda \in (0, 1)$. Recall that q is assumed to be continuous, so q^{σ} is right-dense continuous, and thus delta integrable. Multiplying (2-15) by P^{λ} and delta integrating from $t_1 > \sigma(t_0)$ to t gives

$$\begin{aligned}
&\int_{t_1}^{t} q^{\sigma}(r) P^{\lambda}(r) \Delta r \\
&\le -\int_{t_1}^{t} P^{\lambda}(r) w^{\Delta}(r) \Delta r - \int_{t_1}^{t} p(r) P^{\lambda}(r) (w^{\sigma})^2(r) \Delta r \\
&\overset{\text{parts}}{=} -P^{\lambda}(t) w(t) + P^{\lambda}(t_1) w(t_1) + \int_{t_1}^{t} (P^{\lambda})^{\Delta}(r) w^{\sigma}(r) \Delta r - \int_{t_1}^{t} p(r) P^{\lambda}(r) (w^{\sigma})^2(r) \Delta r \\
&\le -P^{\lambda}(t) w(t) + P^{\lambda}(t_1) w(t_1) \\
&\qquad + \int_{t_1}^{t} \lambda p(r) P^{\lambda - 1}(r) w^{\sigma}(r) \Delta r - \int_{t_1}^{t} p(r) P^{\lambda}(r) (w^{\sigma})^2(r) \Delta r \\
&= -P^{\lambda}(t) w(t) + P^{\lambda}(t_1) w(t_1) + \int_{t_1}^{t} p(r) P^{\lambda - 2}(r) \left[P(r) w^{\sigma}(r) \big(\lambda - P(r) w^{\sigma}(r) \big) \right] \Delta r.
\end{aligned}$$

Since by (2-16) we have

$$0 < P(t)w^\sigma(t) \le P(t)w(t) < 1, \quad t \in [t_0, \infty)_{\mathbb{T}}, \tag{2-17}$$

there exists a positive real number k such that

$$\left| P(r)w^\sigma(r)\big(\lambda - P(r)w^\sigma(r)\big) \right| < k.$$

As a result we have $\lim_{t\to\infty} -P^\lambda(t)w(t) = 0$ by (2-16) for $0 < \lambda < 1$, and

$$\left| \int_{t_1}^{t} p(r)P^{\lambda-2}(r)\left[P(r)w^\sigma(r)\big(\lambda - P(r)w^\sigma(r)\big)\right]\Delta r \right| < k \int_{t_1}^{\infty} p(r)P^{\lambda-2}(r)\Delta r \overset{(2\text{-}9)}{\le} k\frac{(1+\varepsilon)^{2-\lambda}}{1-\lambda}P^{\lambda-1}(t_1)$$

for all $t \in [t_1, \infty)_{\mathbb{T}}$. Therefore we get $\int_{t_1}^{\infty} q^\sigma(r)P^\lambda(r)\Delta r < \infty$, in contradiction with (2-13). □

Due to (2-5) and the establishment of Lemma 2.5, we will henceforth restrict our analysis to the case where

$$\int_{t_0}^{\infty} p(r)\Delta r = \infty, \qquad \int_{t_1}^{\infty} q^\sigma(r)P^\lambda(r)\Delta r < \infty \quad \text{for } \lambda < 1,\ t_1 > \sigma(t_0). \tag{2-18}$$

We also adopt the following notation. Set

$$g(t,\lambda) := \begin{cases} P^{1-\lambda}(t)\int_t^\infty q^\sigma(r)P^\lambda(r)\Delta r & \text{if } \lambda < 1, \\ P^{1-\lambda}(t)\int_{t_0}^t q^\sigma(r)P^\lambda(r)\Delta r & \text{if } \lambda > 1. \end{cases}$$

Then take

$$g_*(\lambda) := \liminf_{t\to\infty} g(t,\lambda) \quad \text{and} \quad g^*(\lambda) := \limsup_{t\to\infty} g(t,\lambda).$$

Lemma 2.6. *Assume that (2-18) holds, that P is given by (2-6), and that (2-7) holds. If (x, y) is a nonoscillatory solution of the system (1-3), then*

$$\liminf_{t\to\infty} w(t)P(t) \ge \tfrac{1}{2}\Big(1 - \sqrt{1 - 4g_*(0)}\Big), \tag{2-19}$$

$$\limsup_{t\to\infty} w(t)P(t) \le \tfrac{1}{2}\Big(1 + \sqrt{1 - 4g_*(2)}\Big), \tag{2-20}$$

where again $w := y/x$.

Proof. By (2-16), we can introduce the constants

$$r := \liminf_{t\to\infty} w(t)P(t), \qquad R := \limsup_{t\to\infty} w(t)P(t), \tag{2-21}$$

and by (2-18), we must have

$$\lim_{t\to\infty} w(t) = 0. \tag{2-22}$$

From (2-14) we have $w^\Delta \le -q^\sigma - pww^\sigma$; delta integrate this from t to ∞, use (2-22), and multiply by P to see that

$$w(t)P(t) \ge P(t)\int_t^\infty q^\sigma(\tau)\Delta\tau + P(t)\int_t^\infty p(\tau)w(\tau)w^\sigma(\tau)\Delta\tau \tag{2-23}$$

holds for $t \in [t_1, \infty)_{\mathbb{T}}$. From (2-21) this yields

$$r \ge g_*(0). \tag{2-24}$$

This time multiply (2-15) by P^2 and delta integrate from t_1 to t to get

$$\begin{aligned}
&\int_{t_1}^t q^\sigma(\tau)P^2(\tau)\Delta\tau \\
&\le -\int_{t_1}^t P^2(\tau)w^\Delta(\tau)\Delta\tau - \int_{t_1}^t p(\tau)P^2(\tau)\big(w^\sigma\big)^2(\tau)\Delta\tau \\
&= -P^2(t)w(t) + P^2(t_1)w(t_1) + \int_{t_1}^t (P^2)^\Delta(\tau)w^\sigma(\tau)\Delta\tau - \int_{t_1}^t p(\tau)P^2(\tau)(w^\sigma)^2(\tau)\Delta\tau \\
&= -P^2(t)w(t) + P^2(t_1)w(t_1) \\
&\qquad + \int_{t_1}^t \mu(\tau)p^2(\tau)w^\sigma(\tau)\Delta\tau + \int_{t_1}^t p(\tau)P(\tau)w^\sigma(\tau)\big[2 - P(\tau)w^\sigma(\tau)\big]\Delta\tau
\end{aligned}$$

for $t \in [t_1, \infty)_{\mathbb{T}}$, which leads to

$$\begin{aligned}
w(t)P(t) \le -P^{-1}(t)&\int_{t_1}^t q^\sigma(\tau)P^2(\tau)\Delta\tau \\
&+ P^{-1}(t)\int_{t_1}^t \mu(\tau)p^2(\tau)w^\sigma(\tau)\Delta\tau + P^{-1}(t)P^2(t_1)w(t_1) \\
&+ P^{-1}(t)\int_{t_1}^t p(\tau)P(\tau)w^\sigma(\tau)\big[2 - P(\tau)w^\sigma(\tau)\big]\Delta\tau.
\end{aligned} \tag{2-25}$$

Using (2-17), we obtain $0 < (1 - Pw^\sigma)^2$, leading to $Pw^\sigma[2 - Pw^\sigma] < 1$. Thus, for large $t \in \mathbb{T}$,

$$P^{-1}(t)\int_{t_1}^t p(\tau)P(\tau)w^\sigma(\tau)\big[2 - P(\tau)w^\sigma(\tau)\big]\Delta\tau \le 1.$$

Applying l'Hôpital's rule [Bohner and Peterson 2001, Theorem 1.120], (2-17) again, and (2-7) we have

$$0 \le \lim_{t\to\infty} \frac{\int_{t_1}^t \mu(\tau)p^2(\tau)w^\sigma(\tau)\Delta\tau}{P(t)} = \lim_{t\to\infty} \mu(t)p(t)w^\sigma(t) \le \lim_{t\to\infty} \frac{\mu(t)p(t)}{P(t)} = 0.$$

Altogether then, inequality (2-25) implies that

$$R \le 1 - g_*(2). \tag{2-26}$$

If $g_*(0) = 0 = g_*(2)$, then estimates (2-19) and (2-20) follow directly from (2-24) and (2-26), respectively. Thus we pick a real number $\varepsilon \in \big(0, \min\{g_*(0), g_*(2)\}\big)$ and $t_2 \in [t_1, \infty)_{\mathbb{T}}$ such that for $t \in [t_2, \infty)_{\mathbb{T}}$,

$$r - \varepsilon < w(t)P(t) < R + \varepsilon, \qquad w(t)P(t) \ge P(t)\int_t^\infty q^\sigma(\tau)\Delta\tau > g_*(0) - \varepsilon,$$
$$P^{-1}(t)\int_{t_0}^t q^\sigma(\tau)P^2(\tau)\Delta\tau > g_*(2) - \varepsilon.$$

From (2-23) and l'Hôpital's rule we have $w(t)P(t) \ge g_*(0) - \varepsilon + (r-\varepsilon)^2$ for $t \in [t_2, \infty)_{\mathbb{T}}$, Multiply (2-14) by P^2 and delta integrate from t_1 to t to see that this leads to

$$\begin{aligned} w(t)P(t) \le {} & -P^{-1}(t)\int_{t_1}^t q^\sigma(\tau)P^2(\tau)\Delta\tau \\ & + P^{-1}(t)\int_{t_1}^t \mu(\tau)p^2(\tau)w^\sigma(\tau)\Delta\tau + P^{-1}(t)P^2(t_1)w(t_1) \\ & + P^{-1}(t)\int_{t_1}^t p(\tau)P(\tau)w^\sigma(\tau)\,[2 - w(\tau)P(\tau)]\,\Delta\tau. \end{aligned} \tag{2-27}$$

From (2-27) we have, for $t \in [t_2, \infty)_{\mathbb{T}}$,

$$w(t)P(t) \le \frac{P^2(t_1)w(t_1) + \int_{t_1}^t \mu(\tau)p^2(\tau)w^\sigma(\tau)\Delta\tau}{P(t)} - g_*(2) + \varepsilon + (R+\varepsilon)(2-R-\varepsilon),$$

since $w^\sigma P \le wP < 1$. These two inequalities lead to

$$r \ge g_*(0) + r^2, \qquad R \le R(2-R) - g_*(2). \tag{2-28}$$

Consequently we have $r \ge \frac{1}{2}\big(1 - \sqrt{1 - 4g_*(0)}\,\big)$ and $R \le \frac{1}{2}\big(1 + \sqrt{1 - 4g_*(2)}\,\big)$, and the lemma is proven. □

3. Main oscillation results

We use the lemmas obtained previously to prove our main results.

Theorem 3.1. *Assume that (2-18) holds, that P is given by (2-6), and that (2-7) holds. If*

$$g_*(0) = \liminf_{t\to\infty} P(t)\int_t^\infty q^\sigma(\tau)\Delta\tau > \frac{1}{4} \tag{3-1}$$

or

$$g_*(2) = \liminf_{t\to\infty} \frac{1}{P(t)}\int_{t_0}^t q^\sigma(\tau)P^2(\tau)\Delta\tau > \frac{1}{4}, \tag{3-2}$$

then every solution of the system (1-3) is oscillatory.

Proof. Suppose to the contrary that (x, y) is a nonoscillatory solution of (1-3) with $x(t) > 0$ for $t \in [t_0, \infty)_{\mathbb{T}}$. Let

$$r := \liminf_{t\to\infty} w(t)P(t), \qquad R := \limsup_{t\to\infty} w(t)P(t),$$

where $w = y/x$. By Lemma 2.6 and its proof (in particular (2-28)) and simple calculus, we have

$$g_*(0) \le r - r^2 \le \tfrac{1}{4} \quad \text{and} \quad g_*(2) \le R - R^2 \le \tfrac{1}{4},$$

in contradiction with both (3-1) and (3-2). □

Theorem 3.2. *Assume that* (2-18) *holds, that P is given by* (2-6), *and that* (2-7) *holds. Let $g_*(2) \le \frac{1}{4}$, and assume there exists a real number $\lambda \in [0, 1)$ such that*

$$g^*(\lambda) > \frac{\lambda^2}{4(1-\lambda)} + \frac{1}{2}\left(1 + \sqrt{1 - 4g_*(2)}\right). \tag{3-3}$$

Then every solution of the system (1-3) *is oscillatory.*

Proof. Suppose to the contrary that (x, y) is a nonoscillatory solution of (1-3) with $x(t) > 0$ for $t \in [t_0, \infty)_{\mathbb{T}}$. By (2-15) we have

$$q^\sigma(t) \le -w^\Delta(t) - p(t)(w^\sigma)^2(t), \quad t \in [t_0, \infty)_{\mathbb{T}},$$

where $w = y/x$; multiply this by P^λ and delta integrate from t to infinity to get

$$\begin{aligned}
&\int_t^\infty q^\sigma(\tau)P^\lambda(\tau)\Delta\tau \\
&\quad \le -\int_t^\infty w^\Delta(\tau)P^\lambda(\tau)\Delta\tau - \int_t^\infty p(\tau)(w^\sigma)^2(\tau)P^\lambda(\tau)\Delta\tau \\
&\quad = P^\lambda(t)w(t) + \int_t^\infty (P^\lambda)^\Delta(\tau)w^\sigma(\tau)\Delta\tau - \int_t^\infty p(\tau)P^\lambda(\tau)(w^\sigma)^2(\tau)\Delta\tau \\
&\quad = P^\lambda(t)w(t) + \frac{1}{4}\int_t^\infty \frac{\left((P^\lambda)^\Delta\right)^2(\tau)}{p(\tau)P^\lambda(\tau)}\Delta\tau \\
&\qquad\qquad - \int_t^\infty \left(\sqrt{p(\tau)}P^{\lambda/2}(\tau)w^\sigma(\tau) - \frac{(P^\lambda)^\Delta(\tau)}{2\sqrt{p(\tau)}P^{\lambda/2}(\tau)}\right)^2\Delta\tau \\
&\quad \le P^\lambda(t)w(t) + \frac{1}{4}\int_t^\infty \frac{\left((P^\lambda)^\Delta\right)^2(\tau)}{p(\tau)P^\lambda(\tau)}\Delta\tau.
\end{aligned}$$

It follows that

$$P^{1-\lambda}(t)\int_t^\infty q^\sigma(\tau)P^\lambda(\tau)\Delta\tau < P(t)w(t) + \frac{P^{1-\lambda}(t)}{4}\int_t^\infty \frac{\left((P^\lambda)^\Delta\right)^2(\tau)}{p(\tau)P^\lambda(\tau)}\Delta\tau. \tag{3-4}$$

By (2-8), (2-20), and (3-4),

$$g^*(\lambda) \le \tfrac{1}{2}\big(1+\sqrt{1-4g_*(2)}\big) + \frac{\lambda^2}{4(1-\lambda)},$$

in contradiction with (3-3). □

Corollary 3.3. *Assume that (2-18) holds, that P is given by (2-6), and that (2-7) holds. If $g_*(2) \le \frac{1}{4}$ and $g^*(0) > \frac{1}{2}\big(1+\sqrt{1-4g_*(2)}\big)$, then every solution of the system (1-3) is oscillatory.*

Theorem 3.4. *Assume that (2-18) holds, that P is given by (2-6), and that (2-7) holds. Let $g_*(0), g_*(2) \le \frac{1}{4}$, and assume there exists a real number $\lambda \in [0, 1)$ such that*

$$g_*(0) > \frac{\lambda(2-\lambda)}{4} \quad \text{and} \quad g^*(\lambda) > \frac{g_*(0)}{1-\lambda} + \tfrac{1}{2}\big(\sqrt{1-4g_*(0)}+\sqrt{1-4g_*(2)}\big). \tag{3-5}$$

Then every solution of the system (1-3) is oscillatory.

Proof. Suppose to the contrary that (x, y) is a nonoscillatory solution of (1-3) with $x(t) > 0$ for $t \in [t_0, \infty)_{\mathbb{T}}$. Set

$$r = \liminf_{t\to\infty} w(t)P(t) \qquad \text{and } R = \limsup_{t\to\infty} w(t)P(t),$$

where $w = y/x$. By (2-19) and (2-20),

$$r \ge m := \tfrac{1}{2}\big(1-\sqrt{1-4g_*(0)}\big) \quad \text{and} \quad R \le M := \tfrac{1}{2}\big(1+\sqrt{1-4g_*(2)}\big). \tag{3-6}$$

Using this and the first inequality in (3-5) we find that $m > \lambda/2$, whence given $\varepsilon \in (0, m-\lambda/2)$, there exists a $t_1 \in [t_0, \infty)_{\mathbb{T}}$ such that

$$m-\varepsilon < w(t)P(t) < M+\varepsilon, \quad t \in [t_1, \infty)_{\mathbb{T}}. \tag{3-7}$$

Similar to what we did before (bottom of page 8), we multiply (2-15) by P^λ and delta integrate from t to infinity to get

$$\int_t^\infty q^\sigma(\tau)P^\lambda(\tau)\Delta\tau \\ \le w(t)P^\lambda(t) + \int_t^\infty p(\tau)P^{\lambda-2}(\tau)\big[\lambda w^\sigma(\tau)P(\tau) - \big(P(\tau)w^\sigma(\tau)\big)^2\big]\Delta\tau;$$

this leads to

$$P^{1-\lambda}(t)\int_t^\infty q^\sigma(\tau)P^\lambda(\tau)\Delta\tau \\ \le w(t)P(t) + P^{1-\lambda}(t)\int_t^\infty p(\tau)P^{\lambda-2}(\tau)\big[\lambda w^\sigma(\tau)P(\tau) - \big(P(\tau)w^\sigma(\tau)\big)^2\big]\Delta\tau.$$

Since the function $f(x) := \lambda x - x^2$ is decreasing over the real interval $[\lambda/2, \infty)$, it follows from the preceding inequality together with (3-7) and Lemma 2.4 that

$$\begin{aligned} P^{1-\lambda}(t)\int_t^\infty q^\sigma(\tau)P^\lambda(\tau)\Delta\tau \\ &< M+\varepsilon+(m-\varepsilon)(\lambda-m+\varepsilon)P^{1-\lambda}(t)\int_t^\infty p(\tau)P^{\lambda-2}(\tau)\Delta\tau \\ &< M+\varepsilon+\frac{(m-\varepsilon)(\lambda-m+\varepsilon)(1+\varepsilon)^{2-\lambda}}{1-\lambda}. \end{aligned}$$

This in tandem with (3-6) yields

$$g^*(\lambda) \le M + \frac{m(\lambda-m)}{1-\lambda} = \frac{g_*(0)}{1-\lambda} + \tfrac{1}{2}\big(\sqrt{1-4g_*(0)} + \sqrt{1-4g_*(2)}\big),$$

in contradiction with the second inequality in (3-5). □

Corollary 3.5. *Assume that* (2-18) *holds, that* P *is given by* (2-6), *and that* (2-7) *holds. Let* $0 < g^*(0) \le \frac{1}{4}$ *and* $g_*(2) \le \frac{1}{4}$. *If*

$$g^*(0) > g_*(0) + \tfrac{1}{2}\big(\sqrt{1-4g_*(0)} + \sqrt{1-4g_*(2)}\big),$$

then every solution of the system (1-3) *is oscillatory.*

4. Examples

We illustrate Theorem 3.1 with the following examples.

Example 4.1. Let $\mathbb{T} = \mathbb{R}$ and $\varepsilon > 0$. Then the continuous linear system

$$x'(t) = \frac{1}{2}\big(\varepsilon + \cos^2 t\big)y(t), \qquad y'(t) = -\frac{1}{t^2}x(t) \tag{4-1}$$

is oscillatory on $[1, \infty)$.

Since $p(t) = \frac{1}{2}(\varepsilon + \cos^2 t)$, we have $P(t) = \frac{1}{8}\big(-2-4\varepsilon+t(2+4\varepsilon)-\sin 2+\sin 2t\big)$. Thus

$$g_*(0) = \liminf_{t\to\infty} P(t)\int_t^\infty q(r)dr = \liminf_{t\to\infty} P(t)/t = \tfrac{1}{4}(1+2\varepsilon) > \tfrac{1}{4}.$$

By Theorem 3.1, any solution pair (x, y) oscillates. Let $x(1) = 0$, $x'(1) = 1$. Numerically generated data for the solutions to (4-1) show a decreasing frequency in oscillations as ε goes to 0, as one might expect. The table shows the value of ε and the estimated value of the first zero of x after $t = 1$.

ε	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
next zero of x	100	128	171	244	379	666	1424	4321	27498

The results led us to wonder whether $\frac{1}{4}$ is sharp for $\mathbb{T} = \mathbb{R}$, and to the next example.

Example 4.2. Let $\mathbb{T}=\mathbb{R}$ and let p be a positive constant. The continuous linear system

$$x'(t)=py(t), \qquad y'(t)=-\frac{1}{t^2}x(t), \qquad t\in[1,\infty) \tag{4-2}$$

is nonoscillatory for $0<p\le\frac{1}{4}$ and oscillatory for $p>\frac{1}{4}$.

Since $p(t)\equiv p$ we have $P(t)=p(t-1)$. Thus

$$g_*(0)=\liminf_{t\to\infty}P(t)\int_t^\infty q(\tau)d\tau=\liminf_{t\to\infty}\frac{p(t-1)}{t}=p.$$

By Theorem 3.1 and (3-1), any solution (x,y) of (4-2) oscillates if $p>\frac{1}{4}$. Converting (4-2) to a second-order equation for x, we arrive at a Cauchy–Euler equation with general solution

$$x(t)=t^{(1-\sqrt{1-4p})/2}\big(A+Bt^{\sqrt{1-4p}}\big).$$

From elementary analysis we know that x is nonoscillatory for $p\le\frac{1}{4}$ and oscillatory for $p>\frac{1}{4}$, showing in particular that the $\frac{1}{4}$ in (3-1) is sharp when $\mathbb{T}=\mathbb{R}$.

5. Future directions: half-linear systems

Let $\varphi_p(x)=|x|^{p-2}x$ for $p>1$ be the one-dimensional p-Laplacian, and consider the following half-linear equation

$$\big[u(t)\varphi_r\big(x^\Delta(t)\big)\big]^\nabla+w(t)\varphi_q\big(x(t)\big)=0, \tag{5-1}$$

where $r,q>1$ and $u,w:\mathbb{T}\to\mathbb{R}$ satisfy $u(t)>0$ and $w(t)\ge 0$, respectively. Let $y=u\varphi_r(x^\Delta)$. Then $y^\nabla=-w\varphi_q(x)$ and $x^\Delta=\varphi_r^{-1}(1/u)\varphi_r^{-1}(y)$, which can be generalized to the half-linear, p-Laplacian system

$$x^\Delta(t)=v(t)\varphi_p\big(y(t)\big), \quad y^\nabla(t)=-w(t)\varphi_q\big(x(t)\big), \quad t\in\mathbb{T}. \tag{5-2}$$

Here we assume $v:\mathbb{T}\to\mathbb{R}$ is right-dense continuous with $v>0$ on $[t_0,\infty)_\mathbb{T}$, and $w:\mathbb{T}\to\mathbb{R}$ is continuous with $w\ge 0$ on $[t_0,\infty)_\mathbb{T}$. Future research would take the earlier results shown valid for (1-3) and try to modify them to cover (5-2) and thus (5-1) as well.

References

[Atici and Guseinov 2002] F. M. Atici and G. S. Guseinov, "On Green's functions and positive solutions for boundary value problems on time scales", *J. Comput. Appl. Math.* **141**:1-2 (2002), 75–99. Dynamic equations on time scales. MR 2003d:34064 Zbl 1007.34025

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

[Bohner and Peterson 2003] M. Bohner and A. Peterson (editors), *Advances in dynamic equations on time scales*, Birkhäuser, Boston, 2003. MR 2004d:34003 Zbl 1025.34001

[Hilger 1990] S. Hilger, "Analysis on measure chains—a unified approach to continuous and discrete calculus", *Results Math.* **18**:1-2 (1990), 18–56. MR 91m:26027 Zbl 0722.39001

[Jiang and Tang 2007] J. Jiang and X. Tang, "Oscillation criteria for two-dimensional difference systems of first order linear difference equations", *Comput. Math. Appl.* **54**:6 (2007), 808–818. MR 2008h:39004 Zbl 1128.34018

[Knaster 1928] B. Knaster, "Un théorème sur les fonctions d'ensembles", *Ann. Soc. Polon. Math.* **6** (1928), 133–134.

[Li 2001] W.-T. Li, "Classification schemes for nonoscillatory solutions of two-dimensional nonlinear difference systems", *Comput. Math. Appl.* **42**:3-5 (2001), 341–355. MR 2003a:39009 Zbl 1006.39012

[Lomtatidze and Partsvania 1999] A. Lomtatidze and N. Partsvania, "Oscillation and nonoscillation criteria for two-dimensional systems of first order linear ordinary differential equations", *Georgian Math. J.* **6**:3 (1999), 285–298. MR 2000c:34089 Zbl 0930.34025

Received: 2007-12-30 Accepted: 2008-10-15

andersod@cord.edu *Department of Mathematics and Computer Science, Concordia College, 901 South 8th Street, Moorhead, MN 56562, United States*
http://www.cord.edu/faculty/andersod/

wrhall@cord.edu *1624 King Cove, Grand Forks, ND 58201, United States*

Zero-divisor ideals and realizable zero-divisor graphs

Michael Axtell, Joe Stickles and Wallace Trampbachls

(Communicated by Scott Chapman)

We seek to classify the sets of zero-divisors that form ideals based on their zero-divisor graphs. We offer full classification of these ideals within finite commutative rings with identity. We also provide various results concerning the realizability of a graph as a zero-divisor graph.

1. Definitions and notation

We will begin by introducing the necessary definitions and notation that will be used throughout the paper. In Section 2, we will determine when the zero-divisors form an ideal in a finite commutative ring with identity, and Section 3 partially generalizes these results to the cases where R is infinite or lacks identity. Section 4 is concerned with the realizability of graphs as zero-divisor graphs.

Given a commutative ring R, an element $x \in R$ is a *zero-divisor* if there exists a nonzero $y \in R$ such that $xy = 0$. We denote the set of zero-divisors as $Z(R)$, and the set of nonzero zero-divisors denoted by $Z(R)^*$. For $x \in R$, the *annihilator* of x, denoted $\mathrm{ann}(x)$, is $\{y \in R \mid xy = 0\}$. It can be shown that the annihilator of any element in a ring is an ideal. An element x is *nilpotent* if $x^n = 0$ for some $n \in \mathbb{N}$. The set of all units in R is denoted $U(R)$. If $x, y \in R$ where R is integral domain, we say x and y are *associates* if $x = uy$, where $u \in U(R)$. A ring R is a *local ring* if and only if R has a unique maximal ideal.

For a graph G, we define $V(G)$ and $E(G)$ to be the sets of vertices and edges of G, respectively. Two elements $x, y \in V(G)$ are defined to be *adjacent*, denoted by

MSC2000: 13A99.

Keywords: zero-divisor, graph, commutative ring.

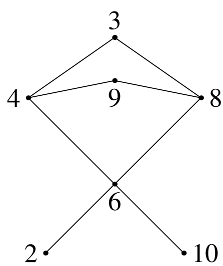
Wallace Trampbachls is a pseudonym which serves to represent a group of student participants at the Wabash Summer Institute in Mathematics in Crawfordsville, Indiana. The first name, Wallace, commemorates Crawfordsville's famous son, Lew Wallace, who authored *Ben Hur*, served as the U.S. Ambassador to the Ottoman Empire, and was a general for the Union Forces in the Civil War. The surname 'Trampbachls' was formed from the first letter of the last name of each student participant. A complete list of the participants is given in the *Acknowledgments*.

x–y, if there exists an edge between them. A *path* between two elements

$$a_1, a_n \in V(G)$$

is an ordered sequence of distinct vertices of G, $\{a_1, a_2, \ldots, a_n\}$, such that a_{i-1}–a_i. The *length* of a path between x and y is the number of edges crossed to get from x to y in the path. The *distance* between $x, y \in G$, denoted $d(x, y)$, is the length of a shortest path between x and y, if such a path exists; otherwise, $d(x, y) = \infty$. For the purposes of this paper, we define $d(x, x) = 0$. The *diameter* of a graph is $\operatorname{diam}(G) = \max\{d(x, y) \mid x, y \in V(G)\}$. An element $x \in V(G)$ is said to be *looped* if there exists an edge from x to itself. A graph G is called *complete bipartite* if there exist disjoint subsets A, B of $V(G)$ such that $A \cup B = V(G)$, $x \not- y$ for any distinct $x, y \in A$ or $x, y \in B$, and x–y for any $x \in A$ and $y \in B$. Finite complete bipartite graphs are denoted as $K^{m,n}$, where $|A| = m$ and $|B| = n$. A graph G is said to be *complete bipartite reducible* if and only if there exists a complete bipartite graph G' such that $V(G') = V(G)$ and $E(G') \subsetneq E(G)$. A graph G is a *star graph* if $G = K^{1,n}$. A graph G is said to be *star-shaped reducible* if and only if there exists a $g \in V(G)$ such that g is adjacent to all other vertices in G and g is looped. More information about graph theory may be found in [Wilson 1972]. We define the *zero-divisor graph* of R, denoted $\Gamma(R)$, as follows: $x \in V(\Gamma(R))$ if and only if $x \in Z(R)^*$, and x–y if and only if $xy = 0$. We will allow loops in $\Gamma(R)$, which is a change from other definitions of zero-divisor graphs as in [Anderson and Livingston 1999; Axtell et al. 2006; Lucas 2006; Redmond 2007].

As an illustration of zero-divisor graphs, we show $\Gamma(\mathbb{Z}_{12})$:



2. Finite rings with identity

In this section, we will ascertain when $Z(R)$ is an ideal in a finite commutative ring with identity by using $\Gamma(R)$, and we will determine the nature of loops in $\Gamma(R)$. Note that to show $Z(R)$ is an ideal, we need only show it is closed under addition.

The following lemma is well known.

Lemma 2.1. *In a finite commutative ring with identity, every element is either a unit or a zero divisor.*

Lemma 2.2. *Let R be a commutative ring. Given any finite set*

$$\{p_1, p_2, \ldots, p_n\} \subset R,$$

where all p_i are nilpotent, there exists a nonzero $a \in R$ such that $ap_i = 0$ for all $1 \le i \le n$.

Proof. Since p_1 is nilpotent, there exists a minimal m_1 such that $p_1^{m_1} = 0$. Let $a_1 = p_1^{m_1-1}$. Clearly, $a_1 p_1 = 0$, and $a_1 \neq 0$. Inductively, since p_i is nilpotent, and $a_{i-1} \neq 0$, there exists $m_i{-}1$ (possibly zero, in which case, $a_i = a_{i-1}$), such that $a_i = a_{i-1} p_i^{m_i-1} \neq 0$, but $a_i p_i = 0$. Let $a = a_n$. By construction, a annihilates every p_i. □

Theorem 2.3. *Let R be a finite commutative ring with identity. Then the following are equivalent*:

(1) $Z(R)$ *is an ideal*;
(2) $Z(R)$ *is a maximal ideal*;
(3) R *is local*;
(4) *Every $x \in Z(R)$ is nilpotent*;
(5) *There exists $b \in Z(R)$ such that $bZ(R) = 0$, and hence $\Gamma(R)$ is star-shaped reducible.*

Proof. Since R is finite, every element is either a zero-divisor or a unit by Lemma 2.1. Hence, whenever $Z(R)$ is an ideal, it must be maximal, since any ideal that properly contains $Z(R)$ must contain a unit and must therefore contain all of R. Also, whenever $Z(R)$ is a maximal ideal, it must be the only one; for consider another ideal I of R. Then either $I \subsetneq Z(R)$, in which case it is not maximal, or else it contains a unit and is not proper. If R is local, then it has a single maximal ideal M. By Lemma 2.1, every element is either a unit or a zero-divisor. In addition, every nonunit is in M, since M is maximal, and every unit is not in M, so M is the set of zero-divisors. Hence the zero-divisors form an ideal. Thus, (1), (2), and (3) are all equivalent.

(1 $\Rightarrow$ 4) Assume $Z(R)$ is an ideal. Suppose $x \in Z(R)$. Then there exist minimal $i > j > 0$ such that $x^i = x^j$. Then, $x^i - x^j = 0$ implies $x^j(x^{i-j} - 1) = 0$. Thus $x^j = 0$ or $x^{i-j} - 1 \in Z(R)$. In the latter case, since $Z(R)$ is an ideal and $x^{i-j} \in Z(R)$, we get $(x^{i-j} - 1) - x^{i-j} = -1 \in Z(R)$, which implies that $Z(R) = R$, a contradiction. Thus $x^j = 0$.

(4 $\Rightarrow$ 5) Assume every element of $Z(R)$ is nilpotent. If $Z(R) = \{0\}$, then condition (5) holds vacuously. Otherwise, by Lemma 2.2 there exists a $b \in Z(R)$ such that $bZ(R) = 0$.

(5 $\Rightarrow$ 1) Assume $\Gamma(R)$ is star-shaped reducible and let b be the center of $\Gamma(R)$. Let x, y be any two elements of $Z(R)$. Then, $b(x-y) = bx - by = 0$, so $x - y \in Z(R)$. Thus, $Z(R)$ is an ideal. □

Corollary 2.4. *If R is a finite commutative ring with identity and* $\operatorname{diam}\Gamma(R)=3$, *then the zero-divisors do not form an ideal.*

Proof. If the $\operatorname{diam}(\Gamma(R))=3$, then $\Gamma(R)$ is not star-shaped reducible. Thus, by Theorem 2.3, $Z(R)$ is not an ideal. □

Theorem 2.5. *For any commutative ring R, if $Z(R)$ is an ideal, then*

$$\Gamma(R)\neq K^{m,n},\, m,n>1.$$

Proof. Assume $Z(R)$ is an ideal. Suppose $\Gamma(R)=K^{m,n}$, $m,n>1$. Let the partition of $\Gamma(R)$ be $A=\{a_1,a_2,\dots,a_m\}$ and $B=\{b_1,b_2,\dots,b_n\}$. Since $Z(R)$ is an ideal, $a_1+b_1\in Z(R)$. If $a_1+b_1=0$, then $a_1=-b_1$, and hence $b_1b_2=0$, a contradiction. Thus, without loss of generality, we may assume that $a_1+b_1\in A$. So, $a_1+b_1=a_i$ for some $i\geq 2$. Since $\Gamma(R)$ is complete bipartite,

$$0=a_ib_2=(a_1+b_1)b_2=a_1b_2+b_1b_2=b_1b_2,$$

a contradiction. Thus $\Gamma(R)\neq K^{m,n}$, $m,n>1$. □

For the remainder of this section we assume that $\Gamma(R)$ is a star graph $K^{1,n}$ with center a. First, we show that the center element of $\Gamma(R)$ is almost always not looped.

Lemma 2.6. *If $\Gamma(R)=K^{1,n}$, then the center element of the star graph a is looped if and only if*

$$R\cong\mathbb{Z}_4,\ \mathbb{Z}_8,\ \mathbb{Z}_9,\ \mathbb{Z}_2[x]/(x^2),\ \mathbb{Z}_2[x]/(x^3),\ \mathbb{Z}_3[x]/(x^2),\ \textit{or}\ \mathbb{Z}_4[x]/(2x,x^2-2).$$

Proof. (⇒) By [Redmond 2007], observe that for $n=0$, 1, 2, we have star graphs with looped centers:

$$\mathbb{Z}_4,\ \mathbb{Z}_8,\ \mathbb{Z}_9,\ \mathbb{Z}_2[x]/(x^2),\ \mathbb{Z}_2[x]/(x^3),\ \mathbb{Z}_3[x]/(x^2),\ \text{ and }\ \mathbb{Z}_4[x]/(2x,x^2-2),$$

respectively. By [Anderson and Livingston 1999, Corollary 2.6], $|\Gamma(R)|>3$, and $\Gamma(R)=K^{1,n}, n>2$ if and only if $R\cong\mathbb{Z}_2\times F$, where F is a field. This implies that the center element of the star graph is $(1,0)$, and $(1,0)$ is not looped.

(⇐) Trivial. □

The previous lemma is useful for determining whether a given graph is a potential zero-divisor graph: if it is a star graph with more than 3 vertices and the center element is looped, then it cannot be a zero-divisor graph.

The next lemma determines when $Z(R)$ is an ideal if $\Gamma(R)$ is a star graph.

Lemma 2.7. *If $\Gamma(R)=K^{1,n}$, then $Z(R)$ is an ideal if and only if*

$$R\cong\mathbb{Z}_4,\ \mathbb{Z}_8,\ \mathbb{Z}_9,\ \mathbb{Z}_2[x]/(x^2),\mathbb{Z}_2[x]/(x^3),\ \mathbb{Z}_3[x]/(x^2),\ \textit{or}\ \mathbb{Z}_4[x]/(2x,x^2-2).$$

Proof. ($\Rightarrow$) Let a be the center of $\Gamma(R)$. Let $b \in Z(R)^*$. Then $a+b \in Z(R)$. Thus, $0 = a(a+b) = a^2 + ab = a^2$. Thus a is looped. By Lemma 2.6,

$$R \cong \mathbb{Z}_4,\ \mathbb{Z}_8,\ \mathbb{Z}_9,\ \mathbb{Z}_2[x]/(x^2),\ \mathbb{Z}_2[x]/(x^3),\ \mathbb{Z}_3[x]/(x^2),\ \textit{or}\ \mathbb{Z}_4[x]/(2x, x^2-2)$$

($\Leftarrow$) Trivial. □

Theorem 2.8. *Let R be a finite commutative ring with identity such that*

$$\Gamma(R) = K^{1,n}$$

with center a. Then the following are equivalent:

(1) $Z(R)$ *is an ideal*;
(2) $a^2 = 0$;
(3) $R \cong \mathbb{Z}_4, \mathbb{Z}_8, \mathbb{Z}_9, \mathbb{Z}_2[x]/(x^2), \mathbb{Z}_2[x]/(x^3), \mathbb{Z}_3[x]/(x^2)$, *or* $\mathbb{Z}_4[x]/(2x, x^2-2)$.

3. General commutative rings

In this section, we will examine the structure of $Z(R)$ with respect to $\Gamma(R)$ in the general case where R does not necessarily have identity or is infinite. By [Anderson and Livingston 1999, Theorem 2.3], we know that $\operatorname{diam}(\Gamma(R)) \leq 3$ for any zero-divisor graph $\Gamma(R)$. We consider each possible diameter of $\Gamma(R)$ separately.

The diameter 0 and 1 cases have already been investigated thoroughly in [Axtell et al. 2006]. In particular, $\operatorname{diam}(\Gamma(R)) = 0$ if and only if $R \cong \mathbb{Z}_4$ or $\mathbb{Z}_2[x]/(x^2)$. In each case, $Z(R)$ forms an ideal. Also, if $\operatorname{diam}(\Gamma(R)) = 1$, then $Z(R)$ is an ideal if and only if $R \ncong \mathbb{Z}_2 \times \mathbb{Z}_2$.

We now expand on the existing results regarding the diameter 2 case and present new results in the diameter 3 case.

The following lemma adds the reverse direction to Lemma 2.3 in [Axtell et al. 2006]; the proof of the forward direction is taken from the same source.

Lemma 3.1. *Let R be a commutative ring such that* $\operatorname{diam}(\Gamma(R)) = 2$. *Then $Z(R)$ is an ideal if and only if for all $x, y \in Z(R)$, there exists a nonzero z such that $xz = yz = 0$.*

Proof. ($\Rightarrow$) Let $x, y \in Z(R)$. If $x = 0$, $y = 0$, or $x = y$, the choice of z to satisfy the statement is clear. Therefore, assume x and y are distinct and nonzero. Since $\operatorname{diam}(\Gamma(R)) = 2$, whenever $xy \neq 0$, there exists $z \in Z(R)^*$ such that $xz = yz = 0$. Thus, assume $xy = 0$. If $x^2 = 0$, then $z = x$ yields the desired element, and likewise if $y^2 = 0$. Suppose $x^2, y^2 \neq 0$. Let $X' = \{x' \in Z(R)^* \mid xx' = 0\}$ and $Y' = \{y' \in Z(R)^* \mid yy' = 0\}$. Observe that $x \in Y'$ and $y \in X'$, so X' and Y' are nonempty. If $X' \cap Y' \neq \varnothing$, choose $z \in X' \cap Y'$. We show $X' \cap Y' \neq \varnothing$. Consider $x + y$. Clearly $x + y \neq x$ and $x + y \neq y$. Also if $x + y = 0$, then $x^2 = 0$ and we are done. If $x+y \neq 0$, since $Z(R)$ is an ideal and thus a subring, we have $x+y \in Z(R)^*$. As $x^2, y^2 \neq 0$, we see that $x+y \notin X'$ and $x+y \notin Y'$. Because $\operatorname{diam}(\Gamma(R)) = 2$, there exists $w \in X'$ such that the following path exists: $x - w - (x+y)$. Then $0 = w(x+y) = wx + wy = wy$ and so $w \in Y'$. Thus, there exists a nonzero z such that $xz = yz = 0$.

($\Leftarrow$) Let $x, y \in Z(R)$. By hypothesis, there exists $z \in Z(R)^*$ such that $xz = yz = 0$. Thus, $(x+y)z = xz + yz = 0$, and $x + y \in Z(R)$. Therefore $Z(R)$ is an ideal. □

Recall Corollary 2.4, which states that there are no finite rings with identity and zero-divisor graph of diameter 3 where $Z(R)$ forms an ideal. This however does not hold for the infinite case. In [Lucas 2006], an example has been given of an infinite ring R in which $Z(R)$ forms an ideal and $\operatorname{diam}(\Gamma(R)) = 3$. We present what we consider to be a more constructive example.

Before we present this example, some notation, definitions, and lemmas are needed. The following definitions are for an integral domain R. An *irreducible element* p is a nonzero, nonunit element that cannot be divided, that is, if $p = qr$, then q or r is a unit. A *unique factorization domain* is an integral domain in which each nonzero nonunit can be factored uniquely, up to associates, as a product of irreducible elements.

Consider the ring $R = \mathbb{Z}_2[x, y, z_1, z_2, \ldots]$. Note that R is a unique factorization domain. Define the set $A = \{p \in R \mid p \in \mathbb{Z}_2[x, y]$ and p is irreducible with zero constant term$\}$. Notice that there are infinitely many such irreducible polynomials in x and y. Indeed, the polynomials $x, x+y, x+y^2, \ldots$ are all irreducible. To see this, consider $\mathbb{Z}_2[x, y]$ in the equivalent form $(\mathbb{Z}_2[y])[x]$. Since $x + y^n$ has degree 1 in x, it can only be factored into something of the form $(f(y)x + g(y)) \cdot (h(y))$. Since the coefficient of x in $x + y^n$ is 1, we must have $f(y), h(y) = \pm 1$. Thus, one of the factors of $x + y^n$, namely $h(y)$, has to be a unit, and thus, $x + y^n$ is irreducible.

Since $\{z_i\}$ and A are countably infinite, there exists a bijection between them, that is, $z_i \to p_i$. Now consider the ideal $Q = (X_1, X_2)$ where $X_1 = \{z_i z_j \mid i, j \in \mathbb{N}\}$ and $X_2 = \{z_i p_i(x, y) \mid i \in \mathbb{N}\}$.

Lemma 3.2. *$f(x, y, z_{i_1}, \ldots, z_{i_n}) + Q \in Z(R/Q)$ if and only if $f(x, y, z_{i_1}, \ldots, z_{i_n})$ has a zero constant term.*

Proof. ($\Leftarrow$) If $f(x, y, z_{i_1}, \ldots, z_{i_n})$ has a zero constant term, then

$$f(x, y, z_{i_1}, \ldots, z_{i_n}) + Q$$

can be written in the form

$$f_{xy} + z_{i_1} f_1 + \cdots + z_{i_n} f_n + Q,$$

where for every k, f_{xy} and f_k are functions in x and y only. Notice that f_{xy} is either irreducible with zero constant term or can be factored into irreducibles, at least one of which has zero constant term, since $\mathbb{Z}_2[x, y]$ is a unique factorization domain. So, there is a z_j such that $z_j f_{xy} + Q = 0 + Q$. Thus,

$$z_j f(x, y, z_{i_1}, \dots, z_{i_n}) + Q = 0 + Q,$$

and hence $f + Q$ is a zero-divisor in R/Q.

($\Rightarrow$) Consider the contrapositive, and assume f has a nonzero constant term. Thus, $f + Q$ cannot be a zero-divisor, since R is an integral domain and no element of Q has a nonzero constant term. □

Proposition 3.3. *In R/Q, $Z(R/Q)$ is an ideal.*

Proof. Since it suffices to show closure under addition, let $f(x, y, z_{i_1}, \dots, z_{i_n}) + Q$, $g(x, y, z_{j_1}, \dots, z_{j_m}) + Q \in Z(R/Q)$. Then

$$(f + g) + Q = h(x, y, z_{i_1}, \dots, z_{i_n}, z_{j_1}, \dots, z_{j_m}) + Q,$$

where h is a polynomial with a zero constant term since f and g both have zero constant terms by Lemma 3.2. □

Theorem 3.4. *In R/Q,* $\operatorname{diam}(\Gamma(R/Q)) = 3$.

Proof. Consider the polynomials $\bar{x} = x + Q$, $\bar{y} = y + Q \in R/Q$. Clearly $\bar{x}$ and $\bar{y}$ $\in Z(R/Q)$, and $\overline{xy} \neq \bar{0}$. Therefore the $d(\bar{x}, \bar{y}) \geq 2$. Suppose $d(\bar{x}, \bar{y}) = 2$. Then there exists $\bar{g} = g + Q \in R/Q$ such that $\bar{x}\bar{g} = \bar{y}\bar{g} = \bar{0}$. By Lemma 3.2, $\bar{g}$ can be written in the form $g_{xy} + z_{i_1} g_1 + z_{i_2} g_2 + \ldots + z_{1_n} g_n + Q$ for some $n \in \mathbb{N}$. Thus, $\bar{x}\bar{g} = x g_{xy} + x z_{i_1} g_1 + x z_{i_2} g_2 + \cdots + x z_{i_n} g_n + Q$. Clearly $x g_{xy} \notin Q$ unless $g_{xy} \in Q$. Thus $\bar{g} = z_{i_1} g_1 + z_{i_2} g_2 + \cdots + z_{i_n} g_n + Q$. However, by construction, there is a unique $\bar{z}_x$ term such that $\bar{x}\bar{z}_x = \bar{0}$. Therefore, $\bar{g} = \bar{g}_{z_x}$, since for any $\bar{z}_i \neq \bar{z}_x$, we have $\bar{x}\bar{z}_i \neq \bar{0}$. An analogous argument holds for $\bar{y}$. Hence, $\bar{g} = \bar{g}_{z_y}$. Therefore, $\bar{z}_x = \bar{z}_y$, a contradiction, since R is a unique factorization domain, and we have a bijection between the indeterminates and the irreducible polynomials. Therefore, $d(\bar{x}, \bar{y}) = 3$ by [Anderson and Livingston 1999]. Thus, $\operatorname{diam}(\Gamma(R/Q)) = 3$. □

Categorizing infinite graphs of diameter 3 for which $Z(R)$ is an ideal is still unresolved.

4. Realizable zero-divisor graphs

In this section, we will analyze the realizability of graphs as zero-divisor graphs of commutative rings with identity through endpoint and cut vertex analysis. We define an *endpoint* to be a vertex that is adjacent to only one other vertex.

Observe that if Γ is a graph on two vertices, it is realizable as a zero-divisor graph of a commutative ring if both endpoints are looped, as can be seen in $\mathbb{Z}_9$ or

$\mathbb{Z}_3[x]/(x^2)$. A two-vertex graph where neither endpoint is looped can be realized as the graph of $\mathbb{Z}_2 \times \mathbb{Z}_2$. If Γ is a graph on two vertices and only one endpoint is looped, then it is not realizable as a zero-divisor graph, as shown by [Redmond 2007].

Theorem 4.1. *Let G be a graph such that $|G| > 2$. If G has at least one looped endpoint, then G is not realizable as the zero-divisor graph of a commutative ring.*

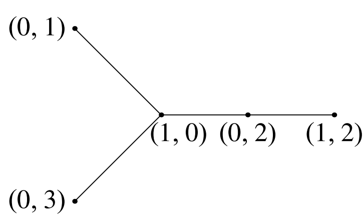
Proof. Assume $G = \Gamma(R)$ for some commutative ring R with identity. Suppose a is a looped endpoint adjacent to a vertex b, and c is a vertex adjacent to b distinct from a in $\Gamma(R)$. Since $a(a+b) = a^2 + ab = 0$, we must have $a+b = a$, $a+b = b$, or $a+b = 0$. If $a+b = a$, then $b = 0$, a contradiction. If $a+b = b$, then $a = 0$, another contradiction. If $a+b = 0$, then $a = -b$ which means any c adjacent to b is adjacent to a, a contradiction. □

A vertex a of a connected graph G is a *cut vertex* if G can be expressed as a union of two subgraphs X and Y such that $E(X) \neq \varnothing$, $E(Y) \neq \varnothing$, $E(X) \cup E(Y) = E(G)$, $V(X) \cup V(Y) = V(G)$, $V(X) \cap V(Y) = \{a\}$, $X \backslash \{a\} \neq \varnothing$, and $Y \backslash \{a\} \neq \varnothing$. In other words, the removal of a cut vertex and its incident edges results in an increase in the number of connected components.

Theorem 4.2. *If $\Gamma(R)$ is partitioned into two subgraphs X and Y with cut vertex a such that $X \backslash \{a\}$ is a complete subgraph, then $I = V(X) \cup \{0\}$ is an ideal.*

Proof. Choose $b \in X \backslash \{a\}$ such that $a - b$. Since $X \backslash \{a\}$ is a complete subgraph and $ab = 0$, we have $bx = 0 = by$ for all $x, y \in V(X) \cup \{0\}$. So, $b(x+y) = 0$, and hence, $x + y \in V(X)$. Similarly, if $r \in R$, we have $b(rx) = r(bx) = 0$, and so $rx \in V(X) \cup \{0\}$. □

The converse of Theorem 4.2 is false. Let $R = \mathbb{Z}_2 \times \mathbb{Z}_4$. Then $(1, 0)$ is a cut vertex. The set $\{(0, 0), (1, 0), (0, 2), (1, 2)\}$ forms an ideal of R; however, their corresponding subgraph is not complete:



Theorem 4.3. *Let R be a commutative ring with identity such that $\Gamma(R)$ is partitioned into two subgraphs X and Y with cut vertex a and $|X| > 2$. If X is a complete subgraph, then every vertex of X is looped.*

Proof. Assume X is a complete subgraph with cut vertex a. Let $b \in X$ such that $b \neq a$. Suppose $b^2 \neq 0$. If $b^2 = b$, then $b(b-1) = 0$, implying $b-1 \in Z(R) \cap V(X)$. Let $c \in V(X)$. Thus, $0 = c(b-1) = cb - c$ implies $c = 0$, a contradiction. If $b^2 \neq b$, then $\operatorname{ann}(b) \subseteq \operatorname{ann}(b^2)$, implying $b^2 \in V(X)$. Thus, $b(b^2) = 0$ since X is complete. So, $b^2(b^2 - b) = 0$, implies $b^2 - b \in Z(R)$. By assumption, $b^2 - b \neq 0$, so $b^2 - b \in X$. Now, $0 = b(b^2 - b) = b^3 - b^2 = -b^2$ yielding $b^2 = 0$.

Now consider the zero-divisor $a+b$. Clearly $a+b \neq a, b$ and since $b(a+b) = 0$, $a+b \in V(X) \cup \{0\}$. Since X is complete, $0 = a(a+b) = a^2$. □

Theorem 4.4. *Let $\Gamma(R)$ have partitions X and Y with cut vertex a. Then $\{0, a\}$ is an ideal.*

Proof. Let $e \in X \backslash \{a\}$ such that $ea = 0$, and let $c \in Y \backslash \{a\}$ such that $ac = 0$. Clearly, $a + a \neq a$. If $a + a = b$ for some $b \in X \backslash \{a\}$, then $c(a+a) = cb = 0$, a contradiction. Similarly, $a + a \notin Y \backslash \{a\}$. Thus, $a + a = 0$. Let $r \in Z(R)$. If $ra \in X \backslash \{a\}$, then $c(ra) = r(ac) = 0$, a contradiction. Similarly, $ra \notin Y \backslash \{a\}$. Thus, $ra \in \{a, 0\}$. □

Theorem 4.5. *If Γ is realizable as a zero-divisor graph of a finite commutative ring with identity, then it is star-shaped reducible, complete bipartite, complete bipartite reducible, or diameter* 3.

Proof. Any finite ring R can be written as $R \cong R_1 \times \cdots \times R_n \times F_1 \times \cdots \times F_m$, where each R_i is local and F_i is a field [Dummit and Foote 2004, p. 752]. If $n+m = 1$, then either R is local or R is a field. If R is local, then zero-divisors form an ideal, and the graph is star-shaped reducible by Theorem 2.3. If R is a field, then $\Gamma(R) = \varnothing$. Now suppose $n+m = 2$. If $R \cong R_1 \times F$, then $\Gamma(R)$ is complete bipartite reducible. If $R \cong F_1 \times F_2$, then $\Gamma(R)$ is complete bipartite. If $R = R_1 \times R_2$, where R_1 and R_2 are local, then let $z \in Z(R_1)^*$, $w \in Z(R_2)^*$. Consider the zero-divisors $z_1 = (z, 1)$ and $z_2 = (1, w)$. The shortest path between z_1 and z_2 must then be of length 3, and hence $\operatorname{diam}(\Gamma(R)) = 3$. If $n + m \geq 3$, $z_1 = (0, 1, 1, \ldots, 1)$ is only attached to $(1, 0, 0, \ldots, 0)$, and $z_2 = (1, 0, 1, \ldots, 1)$ is only attached to $(0, 1, 0, \ldots, 0)$. Since z_1 and z_2 do not have a common annihilator, $\operatorname{diam}(\Gamma(R)) = 3$. □

Corollary 4.6. *A finite graph with no looped vertices is realizable as $\Gamma(R)$ for some commutative ring with identity R if and only if it is the graph of a ring which is a direct product of finite fields.*

Proof. ($\Rightarrow$) If in the decomposition of R, we have that R_1 is local, then by Theorem 2.3, there exists a $k \in R_1$ such that $k^2 = 0$. So, $(k, 0, \ldots, 0)^2 = 0$, and thus $\Gamma(R)$ contains a looped vertex, a contradiction.

($\Leftarrow$) A direct product of fields contains no nonzero nilpotent elements, and hence, $\Gamma(R)$ has no looped vertices. □

Corollary 4.7. *A finite complete bipartite graph* Γ *with partitions* P *and* Q *is realizable as the zero-divisor graph of some commutative ring with identity* R *if and only if* $|P| = p^n - 1$ *and* $|Q| = q^m - 1$ *for some* $m, n \in \mathbb{N}$ *and primes* p, q.

Proof. ($\Rightarrow$) By Theorem 4.5, complete bipartite zero-divisor graphs only arise when $R \cong F_1 \times F_2$. These rings always produce graphs with partitions P and Q such that $|P| = p^n - 1$, and $|Q| = q^m - 1$ for $m, n \in \mathbb{N}$ and primes p, q.

($\Leftarrow$) The ring $R = \mathbb{F}_{p^n} \times \mathbb{F}_{q^m}$ suffices. □

The following two theorems concern the properties of minimal paths in $\Gamma(R)$.

Theorem 4.8. *Let* R *be a commutative ring. If* a–b–c–d *is a minimal path from* a *to* d *in* $\Gamma(R)$, *then* $\operatorname{ann}(a) \subsetneq \operatorname{ann}(c)$. *Furthermore*, $\operatorname{ann}(a)\operatorname{ann}(d) = 0$.

Proof. Since $ad \neq 0$, $\operatorname{ann}(a) \neq \operatorname{ann}(c)$. Suppose there exists $e \in R$ such that $ae = 0$, but $ce \neq 0$. Then $a(ce) = (ae)c = 0$, and $d(ce) = (dc)e = 0$, so $a - ce - d$ is a path of length 2, a contradiction. Thus, $\operatorname{ann}(a) \subsetneq \operatorname{ann}(c)$. Furthermore, suppose there exists $z_a \in \operatorname{ann}(a)$ and $z_d \in \operatorname{ann}(d)$ so that $z_a z_d \neq 0$. Then $a - (z_a z_d) - d$ is also a path of length at most 2, a contradiction. □

Theorem 4.9. *Let* R *be a commutative ring*, *and* $a, d \in Z(R)^*$. *If* a–b–c–d *is a minimal path from* a *to* d, *then* a *and* d *are not nilpotent.*

Proof. Without loss of generality, suppose $a^n = 0$ for some $n \in \mathbb{N}$. Consider the sequence $c, ac, a^2c, a^3c, \ldots, a^nc$. By assumption, $c \neq 0$ and $a^nc = 0$. So, there exists a minimal i such that $a^ic \neq 0$, but $a^{i+1}c = 0$. Thus a^ic is adjacent to both a and d. So $a - a^ic - d$ is a path of length 2, a contradiction. □

5. Acknowledgement

This paper is the result of three summers' worth of undergraduate effort at the Wabash College mathematics REU in collaboration with Professors Axtell and Stickles. The undergraduates involved are Luanne Benson-Lender, University of Evansville (2007), Abigail Bishop, Appalachian State University (2008), Thomas Cuchta, Marshall University (2008), John Holmes, Wabash College (2008), Kathryn Lokken, University of Wisconsin (2008), Michael Martinez, University of Colorado (2006), Oliver Pechenik, Oberlin College (2008), Josiah Roberts, Wabash College (2007), Adam Shull, Valparaiso University (2008), Michael Skalak, Northwestern University (2006), Hilary Smallwood, Fort Lewis College (2008), Drew Swartz, IPFW (2008), and Monty Taylor, University of Tennessee (2007). We would like to thank the referee whose comments proved very useful.

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. 2006] M. Axtell, J. Stickles, and J. Warfel, "Zero-divisor graphs of direct products of commutative rings", *Houston J. Math.* **32**:4 (2006), 985–994 (electronic). MR 2007j:13006 Zbl 1110.13004

[Dummit and Foote 2004] D. S. Dummit and R. M. Foote, *Abstract algebra*, Third ed., John Wiley & Sons Inc., Hoboken, NJ, 2004. MR 2007h:00003 Zbl 1037.00003

[Lucas 2006] T. G. Lucas, "The diameter of a zero divisor graph", *J. Algebra* **301**:1 (2006), 174–193. MR 2007c:13008 Zbl 1109.13006

[Redmond 2007] S. P. Redmond, "On zero-divisor graphs of small finite commutative rings", *Discrete Math.* **307**:9-10 (2007), 1155–1166. MR 2007m:13034 Zbl 1121.13300

[Wilson 1972] R. J. Wilson, *Introduction to graph theory*, Academic Press, New York, 1972. MR 50 #9643 Zbl 0249.05101

Received: 2008-07-10 Revised: 2008-07-16 Accepted: 2008-10-04

maxtell@stthomas.edu *Department of Mathematics, University of St. Thomas, St. Paul, MN 55105, United States*

jstickles@millikin.edu *Department of Mathematics, Millikin University, Decatur, IL 62522, United States*

Wabash Summer Institute in Mathematics, Crawfordsville, IN 47933, United States

Atoms of the relative block monoid

Nicholas Baeth and Justin Hoffmeier

(Communicated by Scott Chapman)

Let G be a finite abelian group with subgroup H and let $\mathcal{F}(G)$ denote the free abelian monoid with basis G. The classical block monoid $\mathcal{B}(G)$ is the collection of sequences in $\mathcal{F}(G)$ whose elements sum to zero. The relative block monoid $\mathcal{B}_H(G)$, defined by Halter-Koch, is the collection of all sequences in $\mathcal{F}(G)$ whose elements sum to an element in H. We use a natural transfer homomorphism $\theta : \mathcal{B}_H(G) \to \mathcal{B}(G/H)$ to enumerate the irreducible elements of $\mathcal{B}_H(G)$ given an enumeration of the irreducible elements of $\mathcal{B}(G/H)$.

1. Introduction

In this paper we will study the so-called block monoid and a generalization called the relative block monoid. The block monoid has been ubiquitous in the literature over the past thirty years and has been used extensively as a tool to study nonunique factorization in certain commutative rings and monoids. The relative block monoid was introduced by Halter-Koch [1992]. Our main goal in this paper is to provide an enumeration of the irreducible elements of the relative block monoid given an enumeration of the irreducible elements of a related block monoid.

In this section we offer a brief description of some central ideas in factorization theory. The quintessential reference for the study of factorization in commutative monoids — in particular block monoids — is [Geroldinger and Halter-Koch 2006, Chapters 5, 6, 7]. In Section 2, we give notations and definitions relevant to studying the relative block monoid. We conclude Section 2 by stating several known results about the relative block monoid. Section 3 provides a means of enumerating the atoms of the relative block monoid $\mathcal{B}_H(G)$ by considering a natural transfer homomorphism $\theta : \mathcal{B}_H(G) \to \mathcal{B}(G/H)$.

For our purposes, a *monoid* is a commutative, cancellative semigroup with identity. We will restrict our attention to reduced monoids, that is, monoids whose set of

MSC2000: 11P70, 20M14.

Keywords: zero-sum sequences, block monoids, finite abelian groups.

This work consists of research done as part of Justin Hoffmeier's Master's thesis at the University of Central Missouri.

units, $H^\times$, contains only the identity element. An element h of a reduced monoid H is said to be *irreducible* or an *atom* if whenever $h = a \cdot b$ with $a, b \in H$, then either $a = 1$ or $b = 1$. We denote the set of atoms of a monoid H by $\mathcal{A}(H)$. If an element $\alpha \in H$ can be written as $\alpha = a_1 \cdots a_k$ with each $a_i \in \mathcal{A}(H)$, this factorization of α is said to have *length* k.

As it is often convenient to study factorization via a surjective map onto a smaller, simpler monoid, we now define transfer homomorphisms. Let H and D be reduced monoids and let $\pi : H \to D$ be a surjective monoid homomorphism. We say that π is a *transfer homomorphism* provided that $\pi^{-1}(1) = \{1\}$ and whenever $\pi(\alpha) = \beta_1\beta_2$ in D, there exist elements α_1 and $\alpha_2 \in H$ such that $\pi(\alpha_1) = \beta_1$, $\pi(\alpha_2) = \beta_2$, and $\alpha = \alpha_1\alpha_2$. It is known that transfer homomorphisms preserve length [Geroldinger and Halter-Koch 2006, Proposition 3.2.3]. That is, if $\pi : H \to D$ is a transfer homomorphism then all questions dealing with lengths of factorizations in H can be studied in D.

2. The relative block monoid

Let G be a finite abelian group written additively and with identity 0. Let $\mathcal{F}(G)$ denote the free abelian monoid with basis G. That is, $\mathcal{F}(G)$ consists of all formal products $g_1^{n_1} \cdots g_k^{n_k}$ with $g_i \in G$ and $n_i \in \mathbb{N}$ with operation given by concatenation. When we write an element $g_1^{n_1} \cdots g_k^{n_k}$ of $\mathcal{B}(G)$ with exponents n_i larger than one, we generally assume that $g_i \neq g_j$ unless $i = j$. We define a monoid homomorphism $\sigma : \mathcal{F}(G) \to G$ by $\sigma(\alpha) = g_1 + \cdots + g_k$ where $\alpha = g_1 g_2 \cdots g_k$. We also use $|\alpha| = n_1 + n_2 + \cdots + n_t$ to denote the length of α in $\mathcal{F}(G)$. We call an element α in $\mathcal{F}(G)$ a *zero-sum sequence* if and only if $\sigma(\alpha) = 0$ in G. If α is a zero-sum sequence and if there does not exist a proper subsequence of α which is also a zero-sum sequence, then we call α a *minimal zero-sum sequence*. The collection of all zero-sum sequences in $\mathcal{F}(G)$, with operation given by concatenation, is called the *block monoid* of G and is denoted $\mathcal{B}(G)$. That is,

$$\mathcal{B}(G) = \{\alpha \in \mathcal{F}(G) \mid \sigma(\alpha) = 0\}.$$

Notice that $\mathcal{B}(G) = \ker(\sigma)$ and that the atoms of $\mathcal{B}(G)$ are simply the nonempty minimal zero-sum sequences. For more general groups, enumerating the atoms of the block monoid is a difficult task. In general, there is no known algorithm to enumerate all atoms of $\mathcal{B}(G)$, although there are some results for special cases of G; see [Geroldinger and Halter-Koch 2006; Ponomarenko 2004]. We will return to this question in Section 3.

When studying zero-sum sequences, the Davenport constant is an important invariant. The *Davenport constant* $\mathsf{D}(G)$ is defined to be the smallest positive integer d such that if $|\alpha| = d$ with $\alpha \in \mathcal{F}(G)$ then there must exist a nonempty subsequence α' of α such that $\sigma(\alpha') = 0$.

Over the past thirty years, several authors have attempted to calculate $\mathsf{D}(G)$ in certain cases, but no general formula is known. What is known about the Davenport constant we summarize in the following theorem [Geroldinger and Halter-Koch 2006]. First we need to define another invariant of a finite abelian group G. If

$$G \cong \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k},$$

with $n_i \mid n_{i+1}$ and $n_i > 1$ for each $1 \leq i < k$, we let

$$\mathsf{d}^*(G) = \sum_{i=1}^{k} (n_i - 1).$$

Theorem 2.1. Let G be a finite abelian group. Then:

(1) $\mathsf{d}^*(G) + 1 \leq \mathsf{D}(G) \leq |G|$;

(2) If G is a cyclic group of order n, then $\mathsf{D}(G) = n$.

We now introduce a somewhat larger submonoid of $\mathcal{F}(G)$, first defined by Halter-Koch [1992]. Let G be a finite abelian group and let H be a subgroup of G. We call an element $\alpha \in \mathcal{F}(G)$ an *H-sum sequence* if $\sigma(\alpha) \in H$. If α is an H-sum sequence and if there does not exist a proper subsequence of an α which is also an H-sum sequence, then α is said to be a *minimal H-sum sequence*. We call the collection of all H-sequences, the *block monoid of G relative to H* and denote it by $\mathcal{B}_H(G)$. Note that if $H = \{0\}$, the H-sum sequences are precisely the zero-sum sequences and hence $\mathcal{B}_H(G) = \mathcal{B}(G)$. In the other extreme case, if $H = G$, then $\mathcal{B}_H(G) = \mathcal{F}(G)$.

As we are now concerned with H-sum sequences, it is natural to define the H-Davenport constant. Let G be a finite abelian group and let H be a subgroup of G. The *H-Davenport constant*, denoted by $\mathsf{D}_H(G)$, is the smallest integer d such that every sequence $\alpha \in \mathcal{F}(G)$ with $|\alpha| \geq d$ has a subsequence $\alpha' \neq 1$ with $\sigma(\alpha') \in H$.

The following theorem [Halter-Koch 1992, Proposition 1] lists several known results about the relative block monoid. We are, in particular, interested in parts 2 and 3 of the theorem.

Theorem 2.2. Let G be an abelian group and let H be a subgroup of G.

(1) The embedding $\mathcal{B}_H(G) \hookrightarrow \mathcal{F}(G)$ is a divisor theory with class group (isomorphic to) G/H and every class contains $|H|$ primes, unless $|G| = 2$ and $H = \{0\}$. If $|G| = 2$ and $H = \{0\}$, then obviously $\mathcal{B}_H(G) = \mathcal{B}(G) \cong (\mathbb{N}_0^2, +)$.

(2) The monoid homomorphism $\theta : \mathcal{B}_H(G) \to \mathcal{B}(G/H)$, defined by

$$\theta(g_1 \cdots g_k) = (g_1 + H) \cdots (g_k + H)$$

is a transfer homomorphism.

(3) $\mathsf{D}_H(G) = \sup\{|\alpha| \mid \alpha \text{ is an atoms of } \mathcal{B}_H(G)\} = \mathsf{D}(G/H)$.

Note that in Theorem 2.2, $|H|$ denotes the cardinality of H while $|\sigma|$ denotes the length of σ. The transfer homomorphism θ from Theorem 2.2 will be heavily used in Section 3 to enumerate the atoms of the relative block monoid.

3. Enumerating the atoms of $\mathcal{B}_H(G)$

Define $N(H)$ to be the number of atoms of a monoid H. In this section we investigate $N(\mathcal{B}_H(G))$. Let G be a finite abelian group and let H be a subgroup. Since $\theta : \mathcal{B}_H(G) \to \mathcal{B}(G/H)$, as defined in Theorem 2.2, is a transfer homomorphism, lengths of factorizations of sequences in $\mathcal{B}_H(G)$ can be studied in the somewhat simpler structure $\mathcal{B}(G/H)$. When G is cyclic of order $n \geq 10$, the number of minimal zero-sum sequences in $\mathcal{B}(G)$ of length $k \geq 2n/3$ is $\phi(n)p_k(n)$ where ϕ is Euler's totient function and where $p_k(n)$ denotes the number of partitions of n into k parts [Ponomarenko 2004, Theorem 8]. Note that by recent work of Yuan [2007, Theorem 3.1] and Savchev and Chen [2007, Proposition 10], the inequality $k \geq 2n/3$ can be replaced by $k \geq \lfloor n/2 \rfloor + 2$ (see also [Geroldinger 2009, Corollary 7.9]). In general, there is no known formula for the number of atoms of $\mathcal{B}(G)$. However, given an enumeration of the atoms of $\mathcal{B}(G/H)$ we can calculate $N(\mathcal{B}_H(G))$ exactly, as the following example illustrates.

Example 1. Let G be a finite abelian group with a subgroup H of index 2. We will calculate $N(\mathcal{B}_H(G))$ as a function of $|H|$, the order of H. Write

$$G/H = \{H, g + H\}, \qquad \text{for some } g \in G \backslash H.$$

It is clear that

$$\mathcal{A}(\mathcal{B}(G/H)) = \{H, (g + H)^2\}.$$

From Theorem 2.2 we know that for each atom $\alpha \in \mathcal{B}_H(G)$, either $\alpha \in \theta^{-1}(H)$ or $\alpha \in \theta^{-1}\left((g + H)^2\right)$. In the first case $|\alpha| = 1$ and so $\alpha \in H$. In the second case, $\alpha = xy$ where $x, y \in g + H$, not necessarily distinct. To count the number of elements of this form, note that we are choosing two elements from the $|H|$ elements of the coset $g + H$. That is, there are $\binom{|H|+1}{2}$ elements in the preimage of $(g_1 + H)^2$. Therefore,

$$N(\mathcal{B}_H(G)) = |H| + \binom{|H| + 1}{2} = \frac{1}{2}|H|^2 + \frac{3}{2}|H|.$$

In the previous example, $N(\mathcal{B}_H(G))$ is a polynomial in $|H|$ with rational coefficients. We now give a series of results to establish this fact in general.

Theorem 3.1. Let G be a finite abelian group and let H be a subgroup of G. If $\alpha = \alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n} \in \mathscr{B}(G/H)$ where $\alpha_i \neq \alpha_j$ whenever $i \neq j$ then

$$\left|\theta^{-1}(\alpha)\right| = \prod_{i=1}^{n}\binom{|H|+t_i-1}{t_i}.$$

Proof. Let

$$\alpha = (x_1+H)^{t_1}(x_2+H)^{t_2}\cdots(x_n+H)^{t_n}$$

be a sequence in $\mathscr{B}(G/H)$ where $x_i+H \neq x_j+H$ unless $i \neq j$. Each element of $\theta^{-1}(x_i+H)^{t_i}$ looks like $y_1y_2\cdots y_{t_i}$ where each $y_j \in x_i+H$. We wish to count the number of such elements in $\mathscr{F}(G)$. Since $\left|\theta^{-1}(x_i+H)\right| = |H|$, we have $|H|$ elements from which to choose. Then to find $\left|\theta^{-1}((x_i+H)^{t_i})\right|$, we choose t_i not necessarily distinct elements from x_i+H. Thus,

$$\left|\theta^{-1}((x_i+H)^{t_i})\right| = \binom{|H|+t_i-1}{t_i}.$$

Since each x_i+H is a distinct coset representative, the elements in the preimage of x_i+H are not in the preimage of any other coset. That is,

$$\theta^{-1}(x_i+H)\cap\theta^{-1}(x_j+H) = \varnothing,$$

whenever $i \neq j$. To find $\left|\theta^{-1}(\alpha)\right|$, we simply multiply, which yields

$$\left|\theta^{-1}(\alpha)\right| = \prod_{i=1}^{n}\binom{|H|+t_i-1}{t_i}. \qquad \square$$

Let $\alpha = \alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n} \in \beta(G/H)$. We say that two sequences $\alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n}$ and $\beta_1^{r_1}\beta_2^{r_2}\cdots\beta_n^{r_n} \in \mathscr{F}(G/H)$ are of *similar form* if

(1) $\alpha_i \neq \alpha_j$ when $i \neq j$,
(2) $\beta_k \neq \beta_l$ when $k \neq l$, and
(3) there exists some $\tau \in S_n$ such that $t_i = r_{\tau(i)}$ for all i.

As we see in the following corollary if α and β are sequences of similar form, then

$$\left|\theta^{-1}(\alpha)\right| = \left|\theta^{-1}(\beta)\right|.$$

Corollary 3.2. Let $\alpha = \alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n}$ and $\beta = \beta_1^{r_1}\beta_2^{r_2}\cdots\beta_n^{r_n} \in \mathscr{F}(G/H)$ be of similar form. Then

$$\left|\theta^{-1}(\alpha)\right| = \left|\theta^{-1}(\beta)\right|.$$

Proof. By Theorem 3.1,

$$\left|\theta^{-1}(\alpha)\right| = \prod_{i=1}^{n}\binom{|H|+t_i-1}{t_i} \qquad \text{and} \qquad \left|\theta^{-1}(\beta)\right| = \prod_{i=1}^{n}\binom{|H|+r_i-1}{r_i}.$$

By assumption, there exists a $\tau \in S_n$ such that $t_i = r_{\tau(i)}$ for all i. Thus, after an appropriate reordering, $t_i = r_i$ for all i. Hence,

$$\left|\theta^{-1}(\alpha)\right| = \prod_{i=1}^{n} \binom{|H|+t_i-1}{t_i} = \prod_{i=1}^{n} \binom{|H|+r_i-1}{r_i} = \left|\theta^{-1}(\beta)\right|. \qquad \square$$

In Example 2, we will categorize the atoms of $\mathcal{B}(G/H)$ to make use of this corollary. We now give our main result. A polynomial $f \in \mathbb{Q}[X]$ is called *integer-valued* if $f(\mathbb{Z}) \subseteq \mathbb{Z}$, and we denote $\mathrm{Int}(\mathbb{Z}) \subset \mathbb{Q}[X]$ the set of integer-valued polynomials on $\mathbb{Z}$. It is well-known that the polynomials $\binom{X}{n}$ form a basis of the $\mathbb{Z}$-module $\mathrm{Int}(\mathbb{Z})$ (see [Cahen and Chabert 1997, Proposition I.1.1]).

Theorem 3.3. Let K be a finite abelian group. There exists an integer-valued polynomial $f \in \mathrm{Int}(\mathbb{Z})$ of degree $\deg(f) = \mathsf{D}(K)$ with the following property: if G is a finite abelian group and $H \subseteq G$ a subgroup with $G/H \cong K$, then

$$N(\mathcal{B}_H(G)) = f(|H|).$$

Proof. From Theorem 2.2 every atom of $\mathcal{B}_H(G)$ is in the preimage of an atom from $\mathcal{B}(G/H)$ under the transfer homomorphism $\theta : \mathcal{B}_H(G) \to \mathcal{B}(G/H)$. Let $A_1, A_2, \ldots, A_m$ denote the atoms of $\mathcal{B}(G/H)$. Then

$$N(\mathcal{B}_H(G)) = \left|\theta^{-1}(A_1)\right| + \left|\theta^{-1}(A_2)\right| + \cdots + \left|\theta^{-1}(A_m)\right|$$

since the preimages $\theta^{-1}(A_i)$ are pairwise disjoint. From Theorem 3.1,

$$\left|\theta^{-1}(A_i)\right| = \prod_{i=1}^{n} \binom{|H|+t_i-1}{t_i}$$

where $A_i = \alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n}$. Since $\binom{|H|+t_i-1}{t_i}$ is a polynomial in terms of $|H|$, we know that $\prod_{i=1}^{n}\binom{|H|+t_i-1}{t_i}$ is a polynomial in terms of $|H|$. Thus,

$$N(\mathcal{B}_H(G)) = \left|\theta^{-1}(A_1)\right| + \left|\theta^{-1}(A_2)\right| + \cdots + \left|\theta^{-1}(A_m)\right|$$

is also a polynomial in terms of $|H|$. The definition of the Davenport constant implies that there exists an atom in $\mathcal{B}(G/H)$ with length $\mathsf{D}(G/H) = \mathsf{D}_H(G)$ and that no longer atom exists. Let $A_i = \alpha_1^{t_1}\alpha_2^{t_2}\cdots\alpha_n^{t_n} \in \mathcal{B}(G/H)$ such that $|A| = \mathsf{D}(G/H) = \mathsf{D}_H(G)$. Then

$$t_1 + t_2 + \cdots + t_n = \mathsf{D}_H(G).$$

Since

$$\binom{|H|+t_i-1}{t_i} = \frac{(|H|+t_i-1)(|H|+t_i-2)\cdots|H|}{t_i}$$

is a polynomial in terms of $|H|$ of degree t_i, $\prod_{i=1}^{n} \binom{|H|+t_i-1}{t_i}$ has degree $\mathsf{D}_H(G)$. Since $\left|A_j\right| \le \mathsf{D}_H(G)$ for all j, we have that

$$N(\mathcal{B}_H(G)) = \left|\theta^{-1}(A_1)\right| + \left|\theta^{-1}(A_2)\right| + \cdots + \left|\theta^{-1}(A_m)\right|,$$

which also has degree $\mathsf{D}_H(G)$. □

Remark 1. If $|H| = 1$, then $H = \{0\}$ and so $\mathcal{B}_H(G) = \mathcal{B}(G)$. In this case, $\left|\theta^{-1}(A_i)\right| = 1$ for all i and thus $N(\mathcal{B}_H(G)) = N(\mathcal{B}(G))$.

We conclude with a final example, which illustrates how much larger $\mathcal{A}(\mathcal{B}_H(G))$ is than $\mathcal{A}(\mathcal{B}(G/H))$.

Example 2. We calculate $N(\mathcal{B}_H(G))$ where $G/H \cong \mathbb{Z}/6\mathbb{Z} = \{0, 1, 2, 3, 4, 5\}$. Note that $\mathcal{A}(\mathcal{B}(\mathbb{Z}/6\mathbb{Z}))$ consists of the following twenty elements:

0	1^6	$1^4 2$	$1^3 3$	$1^2 2^2$	$1^2 4$	123
134^2	15	2^3	$2^2 35$	24	25^2	3^2
345	35^3	4^3	$4^2 5^2$	45^4	5^6	

For each sequence $\alpha \in \mathcal{A}(\mathcal{B}(G/H))$, we compute $\left|\theta^{-1}(\alpha)\right|$. Several pairs of atoms have similar forms and thus we can reduce the number of calculations by using Corollary 3.2. By applying Theorem 3.1 we obtain, for example:

$$\left|\theta^{-1}(3^2)\right| = \binom{|H|+1}{2} = \frac{1}{2}|H|^2 + \frac{1}{2}|H|,$$

$$\left|\theta^{-1}(123, 345)\right| = 2\binom{|H|}{1}^3 = 2|H|^3,$$

and

$$\left|\theta^{-1}(1^4 2, 5^4 4)\right| = 2\binom{|H|+3}{4}\binom{|H|}{1} = \frac{1}{12}|H|^5 + \frac{1}{2}|H|^4 + \frac{11}{12}|H|^3 + \frac{1}{2}|H|^2.$$

These and several similar calculations yield

$$N(\mathcal{B}_H(G)) = \frac{1}{360}|H|^6 + \frac{1}{8}|H|^5 + \frac{185}{72}|H|^4 + \frac{63}{8}|H|^3 + \frac{1247}{180}|H|^2 + \frac{5}{2}|H|.$$

Applying this formula to the case when $|H| = 1$, we find $N(\mathcal{B}_H(G)) = 20$. If $|H| = 10$, then $N(\mathcal{B}_H(G)) = 49,565$, illustrating how quickly $\mathcal{A}(\mathcal{B}_H(G))$ grows as a function of $|H|$.

Acknowledgement

The authors wish to thank the anonymous referee for many insightful comments that greatly improved this paper.

References

[Cahen and Chabert 1997] P.-J. Cahen and J.-L. Chabert, *Integer-valued polynomials*, vol. 48, Mathematical Surveys and Monographs, Am. Math. Soc., Providence, RI, 1997. MR 98a:13002

[Geroldinger 2009] A. Geroldinger, "Additive group theory and non-unique factorizations", in *Combinatorial Number Theory and Additive Group Theory*, edited by A. Geroldinger and I. Ruzsa, Birkhäuser, 2009.

[Geroldinger and Halter-Koch 2006] A. Geroldinger and F. Halter-Koch, *Non-unique factorizations*, vol. 278, Pure and Applied Mathematics (Boca Raton), Chapman & Hall/CRC, Boca Raton, FL, 2006. Algebraic, combinatorial and analytic theory. MR 2006k:20001 Zbl 1113.11002

[Halter-Koch 1992] F. Halter-Koch, "Relative block semigroups and their arithmetical applications", *Comment. Math. Univ. Carolin.* **33**:3 (1992), 373–381. MR 94a:11168 Zbl 0769.11038

[Ponomarenko 2004] V. Ponomarenko, "Minimal zero sequences of finite cyclic groups", *Integers* **4** (2004), A24, 6 pp. (electronic). MR 2005m:11024 Zbl 1083.11015

[Savchev and Chen 2007] S. Savchev and F. Chen, "Long zero-free sequences in finite cyclic groups", *Discrete Math.* **307**:22 (2007), 2671–2679. MR 2008m:11050

[Yuan 2007] P. Yuan, "On the index of minimal zero-sum sequences over finite cyclic groups", *J. Combin. Theory Ser. A* **114**:8 (2007), 1545–1551. MR 2008j:11147 Zbl 1128.11011

Received: 2008-09-15 Revised: 2008-11-10 Accepted: 2008-11-12

baeth@ucmo.edu Mathematics and Computer Science, University of Central Missouri, Warrensburg, MO 64093, United States

hoffmeier@ucmo.edu Mathematics and Computer Science, University of Central Missouri, Warrensburg, MO 64093, United States

INVOLVE 2:1(2009)

Computing points of small height for cubic polynomials

Robert L. Benedetto, Benjamin Dickman, Sasha Joseph, Benjamin Krause, Daniel Rubin and Xinwen Zhou

(Communicated by Bjorn Poonen)

Let $\phi \in \mathbb{Q}[z]$ be a polynomial of degree d at least two. The associated canonical height $\hat{h}_\phi$ is a certain real-valued function on $\mathbb{Q}$ that returns zero precisely at preperiodic rational points of ϕ. Morton and Silverman conjectured in 1994 that the number of such points is bounded above by a constant depending only on d. A related conjecture claims that at nonpreperiodic rational points, $\hat{h}_\phi$ is bounded below by a positive constant (depending only on d) times some kind of height of ϕ itself. In this paper, we provide support for these conjectures in the case $d = 3$ by computing the set of small height points for several billion cubic polynomials.

Let $\phi(z) \in \mathbb{Q}[z]$ be a polynomial with rational coefficients. Define $\phi^0(z) = z$, and for every $n \geq 1$, let $\phi^n(z) = \phi \circ \phi^{n-1}(z)$; that is, ϕ^n is the n-th iterate of ϕ under composition. A point x is said to be *periodic* under ϕ if there is an integer $n \geq 1$ such that $\phi^n(x) = x$. In that case, we say x is n-periodic; the smallest such positive integer n is called the *period* of x. More generally, x is *preperiodic* under ϕ if there are integers $n > m \geq 0$ such that $\phi^n(x) = \phi^m(x)$; equivalently, $\phi^m(x)$ is periodic for some $m \geq 0$.

Using the theory of arithmetic heights, Northcott [1950] proved that if the degree of ϕ is at least 2, then ϕ has only finitely many preperiodic points in $\mathbb{Q}$. (In fact, his result applied far more generally, to morphisms of N-dimensional projective space over any number field.) In 1994, motivated by Northcott's result and by analogies to torsion points of elliptic curves (for which uniform bounds were proven by Mazur [1977] over $\mathbb{Q}$ and by Merel [1996] over arbitrary number fields), Morton and Silverman [1994; 1995] proposed a dynamical Uniform Boundedness Conjecture. Their conjecture applied to the same general setting as Northcott's Theorem, but we state it here only for polynomials over $\mathbb{Q}$.

MSC2000: primary 11G50; secondary 11S99, 37F10.

Keywords: canonical height, p-adic dynamics, preperiodic points.

Conjecture 1 [Morton and Silverman 1994]. *For any $d \geq 2$, there is a constant $M = M(d)$ such that no polynomial $\phi \in \mathbb{Q}[z]$ of degree d has more than M rational preperiodic points.*

Thus far only partial results towards Conjecture 1 have been proven. Several authors [Morton and Silverman 1994; 1995; Narkiewicz 1989; Pezda 1994; Zieve 1996] have bounded the period of a rational periodic point in terms of the smallest prime of good reduction (see Definition 1.3). Others [Flynn et al. 1997; Manes 2007; Morton 1992; 1998; Poonen 1998] have proven that polynomials of degree two cannot have rational periodic points of certain periods by studying the set of rational points on an associated dynamical modular curve; see also [Silverman 2007, Section 4.2]. A different method, introduced in [Call and Goldstine 1997] and generalized and sharpened in [Benedetto 2007], gave (still nonuniform) bounds for the number of preperiodic points by taking into account *all* primes, including those of bad reduction.

In a related vein, the canonical height function $\hat{h}_\phi : \mathbb{Q} \to [0, \infty)$ satisfies the functional equation $\hat{h}_\phi(\phi(z)) = d \cdot \hat{h}_\phi(z)$, where $d = \deg \phi$, and it has the property that $\hat{h}_\phi(x) = 0$ if and only if x is a preperiodic point of ϕ; see Section 1. Meanwhile, if we consider ϕ itself as a point in the appropriate moduli space of all polynomials of degree d, we can also define $h(\phi)$ to be the arithmetic height of that point; see [Silverman 2007, Section 4.11]. For example, the height of the quadratic polynomial $\phi(z) = z^2 + m/n$ is $h(\phi) := h(m/n) = \log \max\{|m|, |n|\}$; a corresponding height for cubic polynomials appears in Definition 4.4. Again by analogy with elliptic curves, we have the following conjecture, stating that the canonical height of a nonpreperiodic rational point cannot be too small in comparison to $h(\phi)$; see [Silverman 2007, Conjecture 4.98] for a more general version.

Conjecture 2. *Let $d \geq 2$. Then there is a positive constant $M' = M'(d) > 0$ such that for any polynomial $\phi \in \mathbb{Q}[z]$ of degree d and any point $x \in \mathbb{Q}$ that is not preperiodic for ϕ, we have $\hat{h}_\phi(x) \geq M' h(\phi)$.*

Just as Conjecture 1 says that any preperiodic rational point must land on a repeated value after a bounded number of iterations, Conjecture 2 essentially says that the size of a nonpreperiodic rational point must start to explode within a bounded number of iterations. Some theoretical evidence for Conjecture 2 appears in [Baker 2006; Ingram $\geq$ 2009], and computational evidence when $d = 2$ appears in [Gillette 2004]. The smallest known value of $\hat{h}_\phi(x)/h(\phi)$ for $d = 2$ occurs for $x = \frac{7}{12}$ under $\phi(z) = z^2 - \frac{181}{144}$; the first few iterates are

$$\frac{7}{12} \mapsto \frac{-11}{12} \mapsto \frac{-5}{12} \mapsto \frac{-13}{12} \mapsto \frac{-1}{12} \mapsto \frac{-5}{4} \mapsto \frac{11}{36} \mapsto \frac{-377}{324} \mapsto \frac{2445}{26244} \mapsto \cdots .$$

(This example was found in [Gillette 2004] by a computer search.) The small canonical height ratio $\hat{h}_\phi(7/12)/\log 181 \approx .0066$ makes precise the observation that although the numerators and denominators of the iterates eventually explode in size, it takes several iterations for the explosion to get underway.

In this paper, we investigate *cubic* polynomials with rational coefficients. We describe an algorithm to find preperiodic and small height rational points of such maps, and we present the resulting data, which supports both conjectures. In particular, after checking the fourteen billion cubics with coefficients of smallest height, we found none with more than eleven rational preperiodic points; those with exactly ten or eleven are listed in Table 2. Meanwhile, as regards Conjecture 2, the smallest height ratio $\mathfrak{h}_\phi(x) := \hat{h}_\phi(x)/h(\phi)$ we found was about .00025, for $\phi(z) = -\frac{25}{24}z^3 + \frac{97}{24}z + 1$ and the point $x = -\frac{7}{5}$, with orbit

$$-\frac{7}{5} \mapsto -\frac{9}{5} \mapsto -\frac{1}{5} \mapsto \frac{1}{5} \mapsto \frac{9}{5} \mapsto \frac{11}{5} \mapsto -\frac{6}{5} \mapsto -\frac{41}{20} \mapsto \frac{4323}{2560} \mapsto \cdots.$$

More importantly, although we found quite a few cubics throughout the search with a nonpreperiodic point of height ratio less than .001, only nine (listed in Table 6) gave $\mathfrak{h}_\phi(x) < .0007$, and the minimal one above was found early in the search. Thus, our data suggests that Conjecture 2 is true for cubic polynomials, with $M'(3) = .00025$.

The outline of the paper is as follows. In Section 1 we review heights, canonical heights, and local canonical heights. In Section 2 we state and prove formulas for estimating local canonical heights accurately in the case of polynomials. In Section 3, we discuss filled Julia sets (both complex and nonarchimedean), and in Section 4 we consider cubics specifically. Finally, we describe our search algorithm in Section 5 and present the resulting data in Section 6.

Our exposition does not assume any background in either dynamics or arithmetic heights, but the interested reader is referred to Silverman's text [Silverman 2007]. For more details on nonarchimedean filled Julia sets and local canonical heights, see [Benedetto 2007; Call and Goldstine 1997; Call and Silverman 1993].

1. Canonical heights

Denote by $M_\mathbb{Q}$ the usual set $\{|\cdot|_\infty, |\cdot|_2, |\cdot|_3, |\cdot|_5, \ldots\}$ of absolute values (also called places) of $\mathbb{Q}$, normalized to satisfy the product formula

$$\prod_{v \in M_\mathbb{Q}} |x|_v = 1 \qquad \text{for any nonzero } x \in \mathbb{Q}^\times.$$

(See [Gouvêa 1997, Chapters 2–3] or [Koblitz 1984, Chapter 1], for example, for background on absolute values.) The standard (global) *height function* on $\mathbb{Q}$ is the function $h : \mathbb{Q} \to \mathbb{R}$ given by $h(x) := \log \max\{|m|_\infty, |n|_\infty\}$, if we write $x = m/n$

in lowest terms. Equivalently,

$$h(x) = \sum_{v \in M_{\mathbb{Q}}} \log \max\{1, |x|_v\} \qquad \text{for any } x \in \mathbb{Q}. \tag{1-1}$$

Of course, h extends to the algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$; see [Lang 1983, Section 3.1], [Hindry and Silverman 2000, Section B.2], or [Silverman 2007, Section 3.1]. The height function satisfies two important properties. First, for any polynomial $\phi(z) \in \mathbb{Q}[z]$, there is a constant $C = C(\phi)$ such that

$$\bigl|h(\phi(x)) - d \cdot h(x)\bigr| \le C \qquad \text{for all } x \in \bar{\mathbb{Q}}, \tag{1-2}$$

where $d = \deg \phi$. Second, if we restrict h to $\mathbb{Q}$, then for any bound $B \in \mathbb{R}$,

$$\{x \in \mathbb{Q} : h(x) \le B\} \quad \text{ is a finite set.} \tag{1-3}$$

For any fixed polynomial $\phi \in \mathbb{Q}[z]$ (or more generally, rational function) of degree $d \ge 2$, the *canonical height* function $\hat{h}_\phi : \bar{\mathbb{Q}} \to \mathbb{R}$ for ϕ is given by

$$\hat{h}_\phi(x) := \lim_{n \to \infty} d^{-n} h(\phi^n(x)),$$

and it satisfies the functional equation

$$\hat{h}_\phi(\phi(x)) = d \cdot \hat{h}_\phi(x) \qquad \text{for all } x \in \bar{\mathbb{Q}}. \tag{1-4}$$

(The convergence of the limit and the functional equation follow fairly easily from (1-2).) In addition, there is a constant $C' = C'(\phi)$ such that

$$\bigl|\hat{h}_\phi(x) - h(x)\bigr| \le C' \qquad \text{for all } x \in \bar{\mathbb{Q}}. \tag{1-5}$$

Northcott's Theorem [Northcott 1950] follows because properties (1-3)–(1-5) imply that for any $x \in \mathbb{Q}$ (in fact, for any $x \in \bar{\mathbb{Q}}$), $\hat{h}(x) = 0$ if and only if x is preperiodic under ϕ.

For our computations, we will need to compute $\hat{h}_\phi(x)$ rapidly and accurately. Unfortunately, the constants C and C' in inequalities (1-2) and (1-5) given by the general theory are rather weak and are rarely described explicitly. The goal of Section 2 will be to improve these constants, using local canonical heights.

Definition 1.1. Let K be a field with absolute value v. We denote by $\mathbb{C}_v$ the completion of an algebraic closure of K. The function $\lambda_v : \mathbb{C}_v \to [0, \infty)$ given by

$$\lambda_v(x) := \log \max\{1, |x|_v\}$$

is called the *standard local height* at v. If $\phi(z) \in K[z]$ is a polynomial of degree $d \ge 2$, the associated *local canonical height* is the function $\hat{\lambda}_{v,\phi} : \mathbb{C}_v \to [0, \infty)$ given by

$$\hat{\lambda}_{v,\phi}(x) := \lim_{n \to \infty} d^{-n} \lambda_v\bigl(\phi^n(x)\bigr). \tag{1-6}$$

According to [Call and Goldstine 1997, Theorem 4.2], the limit in (1-6) converges, so that the definition makes sense. It is immediate that $\hat{\lambda}_{v,\phi}$ satisfies the functional equation $\hat{\lambda}_{v,\phi}(\phi(x)) = d \cdot \hat{\lambda}_{v,\phi}(x)$. In addition, it is well known that $\hat{\lambda}_{v,\phi}(x) - \lambda_v(x)$ is bounded independent of $x \in \mathbb{C}_v$; we shall prove a particular bound in Proposition 2.1 below.

Formula (1-6) of Definition 1.1 is specific to polynomials. For a rational function $\phi = f/g$, where $f, g \in K[z]$ are coprime polynomials and

$$\max\{\deg f, \deg g\} = d \geq 2,$$

the correct functional equation for $\hat{\lambda}_{v,\phi}$ is

$$\hat{\lambda}_{v,\phi}(\phi(x)) = d \cdot \hat{\lambda}_{v,\phi}(x) - \log|g(x)|_v.$$

Of course, formula (1-1) may now be written as

$$h(x) = \sum_{v \in M_{\mathbb{Q}}} \lambda_v(x) \qquad \text{for any } x \in \mathbb{Q}.$$

The local canonical heights provide a similar decomposition for $\hat{h}_\phi$, as follows.

Proposition 1.2. *Let $\phi(z) \in \mathbb{Q}[z]$ be a polynomial of degree $d \geq 2$. Then for all $x \in \mathbb{Q}$,*

$$\hat{h}_\phi(x) = \sum_{v \in M_{\mathbb{Q}}} \hat{\lambda}_{v,\phi}(x).$$

Proof. See [Call and Silverman 1993, Theorem 2.3], which applies to arbitrary number fields, with appropriate modifications. □

Often, the local canonical height $\hat{\lambda}_{v,\phi}$ exactly coincides with the standard local height λ_v; this happens precisely at the places of *good reduction* for ϕ. Good reduction of a map ϕ was first defined in [Morton and Silverman 1994]; see also [Benedetto 2007, Definition 2.1]. For polynomials, it is well known [Morton and Silverman 1995, Example 4.2] that those definitions are equivalent to the following.

Definition 1.3. Let K be a field with absolute value v, and let

$$\phi(z) = a_d z^d + \cdots + a_0 \in K[z]$$

be a polynomial of degree $d \geq 2$. We say that ϕ has *good reduction* at v if

(1) v is nonarchimedean,

(2) $|a_i|_v \leq 1$ for all $i = 0, \ldots, d$, and

(3) $|a_d|_v = 1$.

Otherwise, we say ϕ has *bad reduction* at v.

Note that if $K = \mathbb{Q}$ (or more generally, if K is a global field), a polynomial $\phi \in K[z]$ has bad reduction at only finitely many places $v \in M_K$. As claimed above, we have the following result, proven in, for example, [Call and Goldstine 1997, Theorem 2.2].

Proposition 1.4. *Let K be a field with absolute value v, and let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$ with good reduction at v. Then $\hat{\lambda}_{v,\phi} = \lambda_v$.*

For more background on heights and canonical heights, see [Hindry and Silverman 2000, Section B.2], [Lang 1983, Chapter 3], or [Silverman 2007, Chapter 3]; for local canonical heights, see [Call and Goldstine 1997] or [Call and Silverman 1993, Section 2].

2. Computing local canonical heights

Proposition 2.1. *Let K be a field with absolute value v, let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$, and let $\hat{\lambda}_{v,\phi}$ be the associated local canonical height. Write $\phi(z) = a_d z^d + \cdots + a_1 z + a_0 = a_d(z-\alpha_1)\cdots(z-\alpha_d)$, with $a_i \in K$, $a_d \neq 0$, and $\alpha_i \in \mathbb{C}_v$ Let $A = \max\{|\alpha_i|_v : i = 1, \ldots, d\}$ and $B = |a_d|_v^{-1/d}$, and define real constants $c_v, C_v \geq 1$ by*

$$c_v = \max\{1, A, B\} \qquad \text{and} \qquad C_v = \max\{1, |a_0|_v, |a_1|_v, \ldots, |a_d|_v\}$$

if v is nonarchimedean, or

$$c_v = \max\{1, A + B\} \qquad \text{and} \qquad C_v = \max\{1, |a_0|_v + |a_1|_v + \ldots + |a_d|_v\}$$

if v is archimedean. Then for all $x \in \mathbb{C}_v$,

$$\frac{-d \log c_v}{d-1} \leq \hat{\lambda}_{v,\phi}(x) - \lambda_v(x) \leq \frac{\log C_v}{d-1}.$$

Proof. First, we claim that $\lambda_v(\phi(x)) - d\lambda_v(x) \leq \log C_v$ for any $x \in \mathbb{C}_v$. To see this, if $|x|_v \leq 1$, then $|\phi(x)|_v \leq C_v$, and the desired inequality follows. If $|x|_v > 1$ and $|\phi(x)|_v \leq 1$, the inequality holds because $C_v \geq 1$. Finally, if $|x|_v > 1$ and $|\phi(x)|_v > 1$, then the claim follows from the observation that

$$\left|\frac{\phi(x)}{x^d}\right|_v = \left|a_d + a_{d-1}x^{-1} + \cdots + a_0 x^{-d}\right|_v \leq C_v.$$

Next, we claim that $\lambda_v(\phi(x)) - d\lambda_v(x) \geq -d \log c_v$ for any $x \in \mathbb{C}_v$. If $|x|_v \leq c_v$, then $\lambda_v(x) \leq \log c_v$ because $c_v \geq 1$; the desired inequality is therefore immediate from the fact that $\lambda_v(\phi(x)) \geq 0$. If $|x|_v > c_v$, then

$$\lambda_v(\phi(x)) - d\lambda_v(x) = \lambda_v(\phi(x)) - d \log |x|_v \geq \log |\phi(x)|_v - d \log |x|_v,$$

by definition of λ_v and because $|x|_v > c_v \geq 1$. To prove the claim, then, it suffices to show that $|\phi(x)|_v \geq (|x|_v/c_v)^d$ for $|x|_v > c_v$.

If v is nonarchimedean, then $|x-\alpha_i|_v = |x|_v$ for all $i = 1, \ldots, d$, since $|x|_v > A \geq |\alpha_i|_v$. Hence, $|\phi(x)|_v = |a_d|_v |x|_v^d = (|x|_v/B)^d \geq (|x|_v/c_v)^d$. If v is archimedean, then

$$\frac{|x-\alpha_i|_v}{|x|_v} \geq 1 - \frac{|\alpha_i|_v}{|x|_v} \geq 1 - \frac{A}{A+B} = \frac{B}{A+B} \qquad \text{for all } i = 1, \ldots, d.$$

Thus, $|\phi(x)|_v \geq |a_d|_v (B|x|_v/(A+B))^d = (|x|_v/(A+B))^d \geq (|x|_v/c_v)^d$, as claimed.

To complete the proof, we compute

$$\begin{aligned}
\hat{\lambda}_{v,\phi}(x) - \lambda_v(x) &= \lim_{n\to\infty} \frac{1}{d^n} \lambda_v\big(\phi^n(x)\big) - \lambda_v(x) \\
&= \lim_{n\to\infty} \sum_{j=0}^{n-1} \frac{1}{d^j} \Big[\frac{1}{d} \lambda_v\big(\phi^{j+1}(x)\big) - \lambda_v\big(\phi^j(x)\big) \Big] \\
&\geq \lim_{n\to\infty} \sum_{j=0}^{n-1} -\frac{1}{d^j} \log c_v = -\log c_v \sum_{j=0}^{\infty} \frac{1}{d^j} = \frac{-d \log c_v}{d-1}.
\end{aligned}$$

Similarly, $\hat{\lambda}_{v,\phi}(x) - \lambda_v(x) \leq (\log C_v)/(d-1)$. □

Remark 2.2. The proof above is just an explicit version of [Call and Silverman 1993, Theorem 5.3], giving good bounds for 1, $1/z^d$, $\phi(z)$, and $\phi(z)/z^d$ in certain cases — for example, a lower bound for $|\phi(x)/x^d|_v$ when $|x|_v$ is large. These are precisely the four functions $\{s_{ij}\}_{i,j\in\{0,1\}}$ in [Call and Silverman 1993].

Remark 2.3. If v is nonarchimedean, the quantity $A = \max\{|\alpha_i|_v\}$ can be computed directly from the coefficients of ϕ. Specifically,

$$A = \max\{|a_j/a_d|_v^{1/(d-j)} : 0 \leq j \leq d-1\}.$$

This identity is easy to verify by recognizing $(-1)^{d-j} a_j/a_d$ as the $(d-j)$-th symmetric polynomial in the roots $\{\alpha_i\}$; see also [Call and Goldstine 1997, Lemma 5.1].

On the other hand, if v is archimedean and $|x|_v > \sum_{j=0}^{d-1} |a_j/a_d|_v^{1/(d-j)}$, then

$$\begin{aligned}
|a_d x^d|_v = |x|_v \cdot |a_d x^{d-1}|_v &> \sum_{j=0}^{d-1} \left|\frac{a_j}{a_d}\right|_v^{1/(d-j)} \cdot |x^{d-j-1}|_v \cdot |a_d x^j|_v \\
&\geq \sum_{j=0}^{d-1} \left|\frac{a_j}{a_d}\right|_v \cdot |a_d x^j|_v \\
&= \sum_{j=0}^{d-1} |a_j x^j|_v \geq |a_0 + a_1 x + \cdots + a_{d-1} x^{d-1}|_v,
\end{aligned}$$

and hence $\phi(x) \neq 0$. Thus, $A \leq \sum_{j=0}^{d-1} |a_j/a_d|_v^{1/(d-j)}$ if v is archimedean.

Remark 2.4. Proposition 1.4 can be proven as a corollary of Proposition 2.1, because the constants c_v and C_v are both clearly zero if ϕ has good reduction.

The constants c_v and C_v of Proposition 2.1 can sometimes be improved (that is, made smaller) by changing coordinates, and perhaps even leaving the original base field K. The following proposition shows how local canonical heights change under scaling; but it actually applies to any linear fractional coordinate change.

Proposition 2.5. *Let K be a field with absolute value v, let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$, and let $\gamma \in \mathbb{C}_v^\times$. Define $\psi(z) = \gamma\phi(\gamma^{-1}z) \in \mathbb{C}_v[z]$. Then*

$$\hat{\lambda}_{v,\phi}(x) = \hat{\lambda}_{v,\psi}(\gamma x) \qquad \text{for all } x \in \mathbb{C}_v.$$

Proof. By exchanging ϕ and ψ if necessary, we may assume that $|\gamma|_v \geq 1$. For any $x \in \mathbb{C}_v$ and $n \geq 0$, let $y = \phi^n(x)$. Then $0 \leq \lambda_v(\gamma y) - \lambda_v(y) \leq \log|\gamma|_v$, because $\max\{|y|_v, 1\} \leq \max\{|\gamma y|_v, 1\} \leq |\gamma|_v \max\{|y|_v, 1\}$. Thus,

$$\begin{aligned}\hat{\lambda}_{v,\psi}(\gamma x) - \hat{\lambda}_{v,\phi}(x) &= \lim_{n\to\infty} d^{-n}[\lambda_v(\psi^n(\gamma x)) - \lambda_v(\phi^n(x))] \\ &= \lim_{n\to\infty} d^{-n}[\lambda_v(\gamma\phi^n(x)) - \lambda_v(\phi^n(x))] = 0. \qquad \square\end{aligned}$$

Corollary 2.6. *Let K be a field with absolute value v, let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$, and let $\hat{\lambda}_{v,\phi}$ be the associated local canonical height. Let $\gamma \in \mathbb{C}_v^\times$, and define $\psi(z) = \gamma\phi(\gamma^{-1}z) \in \mathbb{C}_v[z]$. Let c_v and C_v be the constants from Proposition 2.1 for ψ. Then for all $x \in \mathbb{C}_v$,*

$$\frac{-d\log c_v}{d-1} \leq \hat{\lambda}_{v,\phi}(x) - \lambda_v(\gamma x) \leq \frac{\log C_v}{d-1}.$$

We can now prove the main result of this section.

Theorem 2.7. *Let $\phi(z) \in \mathbb{Q}[z]$ be a polynomial of degree $d \geq 2$ with lead coefficient $a \in \mathbb{Q}^\times$. Let $e \geq 1$ be a positive integer, let $\gamma = \sqrt[e]{a} \in \bar{\mathbb{Q}}$ be an e-th root of a, and define $\psi(z) = \gamma\phi(\gamma^{-1}z)$. For each $v \in M_{\mathbb{Q}}$ at which ϕ has bad reduction, let c_v and C_v be the associated constants in Proposition 2.1 for $\psi \in \mathbb{C}_v[z]$. Then*

$$-\frac{1}{d^n}\tilde{c}(\phi, e) \leq \hat{h}_\phi(x) - \frac{1}{ed^n}h\Big(a\big(\phi^n(x)\big)^e\Big) \leq \frac{1}{d^n}\tilde{C}(\phi, e),$$

for all $x \in \mathbb{Q}$ and all integers $n \geq 0$, where

$$\tilde{c}(\phi, e) = \frac{d}{d-1}\sum_{v \text{ bad}} \log c_v, \qquad \textit{and} \qquad \tilde{C}(\phi, e) = \frac{1}{d-1}\sum_{v \text{ bad}} \log C_v.$$

Proof. For any prime v of good reduction for ϕ, we have $|a|_v = 1$; therefore $|\gamma|_v = 1$, and $\lambda_v(\gamma y) = \lambda_v(y)$ for all $y \in \mathbb{C}_v$. Hence, by Equation (1-4), Propositions 1.2

and 1.4, and Corollary 2.6, we compute

$$d^n \hat{h}_\phi(x) = \hat{h}_\phi\bigl(\phi^n(x)\bigr) = \sum_{v\in M_{\mathbb{Q}}} \hat{\lambda}_{v,\phi}\bigl(\phi^n(x)\bigr) = \sum_{v \text{ good}} \lambda_v\bigl(\phi^n(x)\bigr) + \sum_{v \text{ bad}} \hat{\lambda}_{v,\phi}\bigl(\phi^n(x)\bigr)$$

$$\geq -\tilde{c}(\phi, e) + \sum_{v\in M_{\mathbb{Q}}} \lambda_v\bigl(\gamma\,\phi^n(x)\bigr) = -\tilde{c}(\phi, e) + \frac{1}{e}\sum_{v\in M_{\mathbb{Q}}} \lambda_v\Bigl(a\bigl(\phi^n(x)\bigr)^e\Bigr),$$

since $e\lambda_v(y) = \lambda_v(y^e)$ for all $y \in \mathbb{C}_v$. The lower bound is now immediate from the summation formula (1-1). The proof of the upper bound is similar. □

Remark 2.8. The point of Theorem 2.7 is to approximate $\hat{h}_\phi(x)$ even more accurately than the naive estimate $d^{-n}h(\phi^n(x))$, by first changing coordinates to make ϕ monic. Of course, that coordinate change may not be defined over $\mathbb{Q}$; fortunately, the expression $a(\phi^n(x))^e$ at the heart of the theorem still lies in $\mathbb{Q}$, and hence its height is easy to compute quickly.

Remark 2.9. By essentially the same proof, Theorem 2.7 also holds (with appropriate modifications) for any global field K in place of $\mathbb{Q}$.

3. Filled Julia sets

The following definition is standard in both complex and nonarchimedean dynamics.

Definition 3.1. Let K be a field with absolute value v, and let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$. The *filled Julia set* $\mathfrak{K}_v$ of ϕ at v is

$$\mathfrak{K}_v := \bigl\{x \in \mathbb{C}_v : \{|\phi^n(x)|_v : n \geq 0\} \text{ is bounded}\bigr\}.$$

Note that $\phi^{-1}(\mathfrak{K}_v) = \mathfrak{K}_v$. Also note that $\mathfrak{K}_v$ can be defined equivalently as the set of $x \in \mathbb{C}_v$ such that $|\phi^n(x)|_v \not\to \infty$ as $n \to \infty$. In addition, the following well known result relates $\mathfrak{K}_v$ to $\hat{\lambda}_{v,\phi}$; the (easy) proof can be found in [Call and Goldstine 1997, Theorem 6.2].

Proposition 3.2. *Let K be a field with absolute value v, and let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$. For any $x \in \mathbb{C}_v$, we have $\hat{\lambda}_{v,\phi}(x) = 0$ if and only if $x \in \mathfrak{K}_v$.*

Because the local canonical height of a polynomial takes on only nonnegative values, Propositions 1.2 and 3.2 imply that any rational preperiodic points must lie in $\mathfrak{K}_v$ at every place v. However, $\mathfrak{K}_v$ is often a complicated fractal set. Thus, the following Lemmas, which specify disks containing $\mathfrak{K}_v$, will be useful. We set some notation: for any $x \in \mathbb{C}_v$ and $r > 0$, we denote the open and closed disks of radius r about x by

$$D(x, r) = \{y \in \mathbb{C}_v : |y - x|_v < r\} \quad\text{and}\quad \bar{D}(x, r) = \{y \in \mathbb{C}_v : |y - x|_v \leq r\}.$$

Lemma 3.3. *Let K be a field with nonarchimedean absolute value v, let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$ and lead coefficient a_d, and let $\mathfrak{K}_v \subseteq \mathbb{C}_v$ be the filled Julia set of ϕ. Define*

$$s_v = \max\{A, |a_d|_v^{-1/(d-1)}\},$$

where $A = \max\{|\alpha|_v : \phi(\alpha) = 0\}$ as in Proposition 2.1. Then $\mathfrak{K}_v \subseteq \bar{D}(0, s_v)$.

Proof. See [Call and Goldstine 1997, Lemma 5.1]. Alternately, it is easy to check directly that if $|x|_v > s_v$, then

$$|\phi(x)|_v = |a_d x^d|_v > |x|_v.$$

It follows that $|\phi^n(x)|_v \to \infty$. □

Lemma 3.4. *Let K be a field with nonarchimedean absolute value v, and let $\phi(z) \in K[z]$ be a polynomial of degree $d \geq 2$ with lead coefficient a_d. Let $\mathfrak{K}_v \subseteq \mathbb{C}_v$ be the filled Julia set of ϕ at v, let $r_v = \sup\{|x - y|_v : x, y \in \mathfrak{K}_v\}$ be the diameter of $\mathfrak{K}_v$, and let $U_0 \subseteq \mathbb{C}_v$ be the intersection of all disks containing $\mathfrak{K}_v$. Then:*

(1) *$U_0 = \bar{D}(x, r_v)$ for any $x \in \mathfrak{K}_v$.*

(2) *There exists $x \in \mathbb{C}_v$ such that $|x|_v = r_v$.*

(3) *$r_v \geq |a_d|_v^{-1/(d-1)}$, with equality if and only if $\mathfrak{K}_v = U_0$.*

(4) *If $r_v > |a_d|_v^{-1/(d-1)}$, let $\alpha \in U_0$, and let $\beta_1, \dots, \beta_d \in \mathbb{C}_v$ be the roots of $\phi(z) = \alpha$. Then $\mathfrak{K}_v \subseteq U_1$, where*

$$U_1 = \bigcup_{i=1}^{d} \bar{D}(\beta_i, |a_d|_v^{-1/d-1}).$$

Proof. Parts (1–3) are simply a rephrasing of [Benedetto 2007, Lemma 2.5].

As for part (4), if

$$r_v = |a_d|_v^{-1/(d-1)},$$

then $\mathfrak{K}_v = U_0$ by part (3), and hence also

$$\phi^{-1}(U_0) = \phi^{-1}(\mathfrak{K}_v) = \mathfrak{K}_v = U_0.$$

In particular, $\beta_i \in U_0$ for all i, and the result follows.

If

$$r_v > |a_d|_v^{-1/(d-1)},$$

Benedetto [2007, Lemma 2.7] says that $\phi^{-1}(U_0)$ is a disjoint union of ℓ strictly smaller disks $V_1, \dots, V_\ell$, each contained in U_0, and each of which maps onto U_0 under ϕ, for some integer $2 \leq \ell \leq d$.

Suppose there is some $x \in \mathfrak{K}_v$ such that

$$|x - \beta_i|_v > |a_d|_v^{-1/(d-1)} \quad \text{for all } i = 1, \dots, d.$$

By part (1), there is some $y \in \mathfrak{K}_v$ such that $|x-y|_v = r_v$. Without loss, $x \in V_1$ and $y \in V_2$; V_1 and V_2 are distinct and in fact disjoint, because each has radius strictly smaller than r_v, and v is nonarchimedean. The disk V_2 must also contain some β_j (without loss, β_d), since $\phi(V_2) = U_0$ by the previous paragraph; hence $|x-\beta_d|_v = r_v$. Thus,

$$|\phi(x)-\alpha|_v = |a_d|_v \cdot |x-\beta_d|_v \prod_{i=1}^{d-1} |x-\beta_i|_v > |a_d|_v \cdot r_v \cdot (|a_d|_v^{-1/(d-1)})^{d-1} = r_v.$$

However,

$$\phi(x) \in \mathfrak{K}_v \subseteq U_0 \quad \text{and} \quad \alpha \in U_0;$$

therefore $|\phi(x)-\alpha|_v \le r_v$. Contradiction. □

Remark 3.5. Lemma 3.4(4) says that $\mathfrak{K}_v$ is contained in a union of at most d disks of radius $|a_d|_v^{-1/(d-1)}$. However, if $d \ge 3$, then at most one of the disks needs to be that large; the rest can be strictly smaller. Still, the weaker statement of Lemma 3.4 above suffices for our purposes.

4. Cubic polynomials

In the study of quadratic polynomial dynamics, it is useful to note that (except in characteristic 2) any such polynomial is conjugate over the base field to a unique one of the form z^2+c. For cubics, it might appear at first glance that a good corresponding form would be z^3+az+b. However, this form is not unique, since z^3+az+b is conjugate to z^3+az-b by $z \mapsto -z$. In addition, it is not even possible to make most cubic polynomials monic by conjugation over $\mathbb{Q}$. More precisely, if ϕ is a cubic with leading coefficient a, and if $\eta(z) = \alpha z + \beta$, then $\eta^{-1} \circ \phi \circ \eta$ has leading coefficient $\alpha^{-2} a$, which can only be 1 if a is a perfect square. Instead of z^3+az+b, then, we propose the following two forms as normal forms when conjugating over a (not necessarily algebraically closed) field of characteristic not equal to three.

Definition 4.1. Let K be a field, and let $\phi \in K[z]$ be a cubic polynomial. We will say that ϕ is *in normal form* if either

$$\phi(z) = az^3 + bz + 1 \tag{4-1}$$

or

$$\phi(z) = az^3 + bz. \tag{4-2}$$

Proposition 4.2. *Let K be a field of characteristic not equal to* 3, *and let $\phi(z) \in K[z]$ be a cubic polynomial. Then there is a degree one polynomial $\eta \in K[z]$ such that $\psi = \eta^{-1} \circ \phi \circ \eta$ is in normal form. Moreover, if another conjugacy $\tilde{\eta}(z)$ also gives a normal form $\tilde{\psi} = \tilde{\eta}^{-1} \circ \phi \circ \tilde{\eta}$, then either $\tilde{\eta} = \eta$ and $\tilde{\psi} = \psi$, or else both* *normal forms* $\psi(z) = az^3 + bz$ *and* $\tilde{\psi}(z) = \tilde{a}z^3 + bz$ *are of the type in* (4-2) *with the same linear term, and the quotient* $\tilde{a}/a$ *of their lead coefficients is the square of an element of* K.

Proof. Write $\phi(z) = az^3 + bz^2 + cz + d \in K[z]$, with $a \neq 0$. Conjugating by $\eta_1(z) = z - b/(3a)$ gives

$$\psi_1(z) := \eta_1^{-1} \circ \phi \circ \eta_1(z) = az^3 + b'z + d'.$$

(Note that $b', d' \in K$ can be computed explicitly in terms of a, b, c, d, but their precise values are not important here.) If $d' = 0$, then we have a normal form of the type in (4-2). Otherwise, conjugating ψ_1 by $\eta_2(z) = d'z$ gives the normal form

$$\eta_2^{-1} \circ \psi_1 \circ \eta_2(z) = a'z^3 + b'z + 1,$$

where $a' = a/(d')^2$.

For the uniqueness, suppose $\phi_1 = \eta^{-1} \circ \phi_2 \circ \eta$, where

$$\eta(z) = \alpha z + \beta, \qquad \phi_1(z) = a_1 z^3 + b_1 z + c_1, \qquad \phi_2(z) = a_2 z^3 + b_2 z + c_2,$$

with $c_1, c_2 \in \{0, 1\}$ and $\alpha a_1 a_2 \neq 0$. Because the z^2-coefficient of $\eta^{-1} \circ \phi_2 \circ \eta(z)$ is $\alpha\beta a_1$, we must have $\beta = 0$. Thus,

$$\phi_2(z) = \alpha^{-1} \phi_1(\alpha z),$$

which means that $c_2 = \alpha c_1$ and $a_2/a_1 = \alpha^2$. If either c_1 or c_2 is 1, then $\alpha = 1$ and $\phi_1 = \phi_2$. Otherwise, we have $c_1 = c_2 = 0$, $b_1 = b_2$, and $a_2/a_1 \in (K^\times)^2$, as claimed. □

Remark 4.3. The cubic $\phi(z) = az^3 + bz$ is self-conjugate under $z \mapsto -z$; that is, $\phi(-z) = -\phi(z)$. (It is not a coincidence that those cubic polynomials admitting nontrivial self-conjugacies are precisely those with the more complicated "$\tilde{a}/a$ is a square" condition in Proposition 4.2; see [Silverman 2007, Example 4.75 and Theorem 4.79].) As a result, $\hat{h}_\phi(-x) = \hat{h}_\phi(x)$ for all $x \in \mathbb{Q}$; and if x is a preperiodic point of ϕ, then so is $-x$.

In addition, the function $-\phi(z) = -az^3 - bz$ satisfies $(-\phi) \circ (-\phi) = \phi \circ \phi$. Thus, $\hat{h}_\phi(x) = \hat{h}_{-\phi}(x)$ for all $x \in \mathbb{Q}$. Moreover, ϕ and $-\phi$ have the same set of preperiodic points, albeit with slightly different arrangements of points into cycles.

The normal forms of Definition 4.1 have two key uses. The first is that they allow us to list a unique (or, in the case of form (4-2), essentially unique) element of each conjugacy class of cubic polynomials over $\mathbb{Q}$ in a systematic way, which is helpful for having a computer algorithm test them one at a time. The second is that the forms provide a description of the moduli space $\mathcal{M}_3$ of all cubic polynomials up to conjugation. This second use is crucial to the very statement of Conjecture 2, because the quantity $h(\phi)$ is defined to be the height of the conjugacy class of ϕ viewed as a point on $\mathcal{M}_3$.

In particular, Proposition 4.2 says that $\mathcal{M}_3$ can be partitioned into two pieces: the first piece is an affine subvariety of $\mathbb{P}^2$, and the second is an affine line. More specifically, the conjugacy class of the polynomial $\phi(z) = az^3 + bz + 1$ corresponds to the point (a, b) in $\{(a, b) \in \mathbb{A}^2 : a \neq 0\}$. To compute heights, then, we should view $\mathbb{A}^2$ as an affine subvariety of $\mathbb{P}^2$, thus declaring $h(\phi)$ to be the height $h([a : b : 1])$ of the point $[a : b : 1]$ in $\mathbb{P}^2$. Meanwhile, the conjugacy class of $az^3 + bz$ over $\bar{\mathbb{Q}}$ is determined solely by b, because $az^3 + bz$ is conjugate to $a'z^3 + bz$ over $\bar{\mathbb{Q}}$. (As noted in [Silverman 2007, Section 4.4 and Remark 4.39], $\mathcal{M}_3$ is the moduli space of $\bar{\mathbb{Q}}$-conjugacy classes of cubic polynomials, not $\mathbb{Q}$-conjugacy classes.) Thus, the $\bar{\mathbb{Q}}$-conjugacy class of $\phi_0(z) = az^3 + bz$ corresponds to the point b in $\mathbb{A}^1$, and the corresponding height is $h(\phi_0) = h([b : 1])$, the height of the point $[b : 1]$ in $\mathbb{P}^1$. We phrase these assignments formally in the following definition.

Definition 4.4. Given $a, b \in \mathbb{Q}$ with $a \neq 0$, define

$$\phi(z) = az^3 + bz + 1 \quad \text{and} \quad \phi_0(z) = az^3 + bz.$$

Write $a = k/m$ and $b = \ell/m$ with $k, \ell, m \in \mathbb{Z}$ and $\gcd(k, \ell, m) = 1$; also write $b = \ell_0/m_0$ with $\gcd(\ell_0, m_0) = 1$. Then we define the *heights* $h(\phi)$, $h(\phi_0)$ of the maps ϕ and ϕ_0 to be

$$h(\phi) := \log\max\{|k|_\infty, |\ell|_\infty, |m|_\infty\} \qquad \text{and} \qquad h(\phi_0) := \log\max\{|\ell_0|_\infty, |m_0|_\infty\}.$$

Note that

$$h(\phi_0) = h(b) = \sum_v \log\max\{1, |b|_v\}, \qquad \text{and} \qquad h(\phi) = \sum_v \log\max\{1, |a|_v, |b|_v\}.$$

Proposition 4.5. *Given a, b, ϕ, ϕ_0 as in Definition 4.4, let $\gamma = \sqrt{a} \in \bar{\mathbb{Q}}$ be a square root of a, and define*

$$\psi(z) = \gamma\,\phi(\gamma^{-1}z) = z^3 + bz + \sqrt{a}, \qquad \textit{and} \qquad \psi_0(z) = \gamma\,\phi_0(\gamma^{-1}z) = z^3 + bz.$$

Let $\tilde{c}(\phi, 2)$, $\tilde{C}(\phi, 2)$, $\tilde{c}(\phi_0, 2)$, and $\tilde{C}(\phi_0, 2)$ be the corresponding constants from Theorem 2.7. Then

$$\tilde{c}(\phi, 2) \le 1.84 \cdot \max\{h(\phi), 1\}, \qquad \tilde{C}(\phi, 2) \le .75 \cdot \max\{h(\phi), 1\},$$
$$\tilde{c}(\phi_0, 2) \le 1.57 \cdot \max\{h(\phi_0), 1\}, \qquad \tilde{C}(\phi_0, 2) \le .75 \cdot h(\phi_0).$$

Proof. Note that

$$\begin{aligned} \log(1 + |a|_\infty^{1/6} + |b|_\infty^{1/2}) &\le \log\bigl(3\max\{1, |a|_\infty^{1/6}, |b|_\infty^{1/2}\}\bigr) \\ &\le \log 3 + \tfrac{1}{2}\log\max\{1, |a|_\infty, |b|_\infty\}. \end{aligned}$$

Thus, if $h(\phi) \geq \log 9$, then by Remark 2.3 and the definition of $\tilde{c}(\phi, 2)$,

$$\begin{aligned}\frac{2}{3}\tilde{c}(\phi, 2) &\leq \log(1 + |a|_\infty^{1/6} + |b|_\infty^{1/2}) + \sum_{v \neq \infty} \log \max\{1, |a|_v^{1/6}, |b|_v^{1/2}\} \\ &\leq \log 3 + \frac{1}{2}\sum_v \log \max\{1, |a|_v, |b|_v\} = \log 3 + \frac{1}{2}h(\phi) \leq h(\phi).\end{aligned}$$

Hence, $\tilde{c}(\phi, 2) < 1.5 \cdot h(\phi)$ if $h(\phi) > \log 9$.

Similarly,

$$\log(1 + |a|_\infty + |b|_\infty) \leq \log 3 + \log \max\{1, |a|_\infty, |b|_\infty\},$$

and therefore

$$2\tilde{C}(\phi, 2) \leq \log 3 + h(\phi) \leq 1.5 \cdot h(\phi)$$

if $h(\phi) \geq \log 9$. The bounds for ϕ_0 can be proven in the same fashion in the case that $h(\phi_0) \geq \log 4$. (That is, $h(b) \geq 4$.)

Finally, there are fifteen choices of $b \in \mathbb{Q}$ for which $h(b) < \log 4$, and 1842 pairs $(a, b) \in \mathbb{Q}$ for which $h(\phi) < \log 9$. By a simple computer computation (working directly from the definitions in Proposition 2.1 and Theorem 2.7, not the estimates of Remark 2.3), one can check that the desired inequalities hold in all cases. □

Remark 4.6. In fact, $\tilde{c}(\phi_0, 2) \leq 1.5 \cdot \max\{h(b), 1\}$ in all but four cases: $b = \pm 2/3$ and $b = \pm 3/2$, which give $h(b) = \log 3$ and $\tilde{c}(\phi_0, 2) = 1.5 \cdot \log(\sqrt{2} + \sqrt{3})$. Similarly, $\tilde{c}(\phi, 2) \leq 1.5 \cdot \max\{h(\phi), 1\}$ in all but 80 cases. The maximum ratio of $1.838\ldots$ is attained twice, when (a, b) is $(-1, 2/3)$ or $(1, -2/3)$. In both cases, $h(\phi) = \log 3$ and $\tilde{c}(\phi_0, 2) = 1.5 \cdot \log((\alpha + 1)\sqrt{3})$, where $\alpha \approx 1.22$ is the unique real root of $3z^3 - 2z - 3$.

The next lemma says that for cubic polynomials in normal form, and for v a p-adic absolute value with $p \neq 3$, the radius s_v from Lemma 3.3 coincides with the radius r_v from Lemma 3.4. Thus, when we search for rational preperiodic points, we are losing no efficiency by searching in $\bar{D}(0, s_v)$ instead of the ostensibly smaller disk U_0.

Lemma 4.7. *Let K be a field with nonarchimedean absolute value v such that $|3|_v = 1$. Let $\phi(z) \in K[z]$ be a cubic polynomial in normal form, and let $\mathfrak{K}_v \subseteq \mathbb{C}_v$ be the filled Julia set of ϕ at v. Let $r_v = \sup\{|x - y|_v : x, y \in \mathfrak{K}_v\}$ be the diameter of $\mathfrak{K}_v$. Then $|x|_v \leq r_v$ for all $x \in \mathfrak{K}_v$.*

Proof. If $\phi(z) = az^3 + bz$, then $\phi(0) = 0$, and therefore $0 \in \mathfrak{K}_v$. The desired conclusion is immediate. Thus, we consider $\phi(z) = az^3 + bz + 1$. Note that the three roots $\alpha, \beta, \gamma \in \mathbb{C}_v$ of the equation $\phi(z) - z = 0$ are fixed by ϕ and hence lie in $\mathfrak{K}_v$.

Without loss, assume $|\alpha|_v \geq |\beta|_v \geq |\gamma|_v$. It suffices to show that $|\alpha - \gamma|_v = |\alpha|_v$; if $x \in \mathfrak{K}_v$, then $|x|_v \leq \max\{|x - \alpha|_v, |\alpha - \gamma|_v\} \leq r_v$, as desired. If $|\alpha|_v > |\gamma|_v$, then

$|\alpha-\gamma|_v=|\alpha|_v$, and we are done. Thus, we may assume $|\alpha|_v=|\beta|_v=|\gamma|_v=|a|_v^{-1/3}$, which implies that $|a|_v \ge |b-1|_v^3$. We may also assume that $|\alpha-\gamma|_v \ge |\alpha-\beta|_v$.

The polynomial $Q(z)=\phi(z+\alpha)-(z+\alpha)$ has roots 0, $\beta-\alpha$, and $\gamma-\alpha$; on the other hand, $Q(z)=az[z^2+3\alpha z+(a^{-1}(b-1)+3\alpha^2)]$ by direct computation. Thus,

$$\bigl(z-(\beta-\alpha)\bigr)\bigl(z-(\gamma-\alpha)\bigr)=z^2+3\alpha z+\bigl(a^{-1}(b-1)+3\alpha^2\bigr). \tag{4-3}$$

Since $|a^{-1}(b-1)|_v \le |a|_v^{-2/3}=|\alpha|_v^2$, the constant term of (4-3) has absolute value at most $|\alpha|_v^2$; meanwhile, the linear coefficient satisfies $|3\alpha|_v=|\alpha|_v$. Thus, either from the Newton polygon or simply by inspection of (4-3), it follows that $|\alpha-\gamma|_v=|\alpha|_v$. □

Remark 4.8. Lemma 4.7 can be false in nonarchimedean fields in which $|3|_v<1$. For example, if $K=\mathbb{Q}_3$ (in which $|3|_3=1/3<1$) and $\phi(z)=-(1/27)z^3+z+1$, then it is not difficult to show that the diameter of the filled Julia set is $3^{-3/2}$. However, $\alpha=3$ is a fixed point, and $|\alpha|_3=1/3>3^{-3/2}$.

At the archimedean place $v=\infty$, we will study not $\mathfrak{K}_\infty$ itself, but rather the simpler set $\mathfrak{K}_\infty\cap\mathbb{R}$, which we will describe in terms of the real fixed points. Note, of course, that any cubic with real coefficients has at least one real fixed point; and if there are exactly two real fixed points, then one must appear with multiplicity two.

Lemma 4.9. *Let $\phi(z)\in\mathbb{R}[z]$ be a cubic polynomial with positive lead coefficient. If ϕ has precisely one real fixed point $\gamma\in\mathbb{R}$, then $\mathfrak{K}_\infty\cap\mathbb{R}=\{\gamma\}$ is a single point.*

Proof. We can write $\phi(z)=z+(z-\gamma)^j\psi(z)$, where $1\le j\le 3$, and where $\psi\in\mathbb{R}[z]$ has positive lead coefficient and no real roots. Thus, there is a positive constant $c>0$ such that $\psi(x)\ge c$ for all $x\in\mathbb{R}$. Given any $x\in\mathbb{R}$ with $x>\gamma$, then, $\phi(x)>x+c(x-\gamma)^j$. It follows that $\phi^n(x)>x+nc(x-\gamma)^j$, and hence $\phi^n(x)\to\infty$ as $n\to\infty$. Similarly, for $x<\gamma$, $\phi^n(x)\to-\infty$ as $n\to\infty$. □

Lemma 4.10. *Let $\phi(z)\in\mathbb{R}[z]$ be a cubic polynomial with positive lead coefficient $a>0$ and at least two distinct fixed points. Denote the fixed points by $\gamma_1,\gamma_2,\gamma_3\in\mathbb{R}$, with $\gamma_1\le\gamma_2\le\gamma_3$. Then $\mathfrak{K}_\infty\cap\mathbb{R}\subseteq[\gamma_1,\gamma_3]$, and*

$$\phi^{-1}([\gamma_1,\gamma_3])\subseteq[\gamma_1,\gamma_1+a^{-1/2}]\cup[\gamma_2-a^{-1/2},\gamma_2+a^{-1/2}]\cup[\gamma_3-a^{-1/2},\gamma_3].$$

Proof. Let $\alpha=\inf(\mathfrak{K}_\infty\cap\mathbb{R})$; then $\alpha\in\mathfrak{K}_\infty\cap\mathbb{R}$, since this set is closed. Therefore, $\phi(\alpha)\ge\alpha$, because $\phi(\mathfrak{K}_\infty\cap\mathbb{R})\subseteq\mathfrak{K}_\infty\cap\mathbb{R}$. On the other hand, if $\phi(\alpha)>\alpha$, then by continuity (and because ϕ has positive lead coefficient), there is some $\alpha'<\alpha$ such that $\phi(\alpha')=\alpha$, contradicting the minimality of α. Thus $\phi(\alpha)=\alpha$, giving $\alpha=\gamma_1$. Similarly, $\sup(\mathfrak{K}_\infty)=\gamma_3$, proving the first statement.

For the second statement, note that

$$\phi(z) = a(z-\gamma_1)(z-\gamma_2)(z-\gamma_3) + z,$$

and consider $x \in \mathbb{R}$ outside all three desired intervals. We will show that

$$\phi(x) \notin [\gamma_1, \gamma_3].$$

If $x > \gamma_3$, then $\phi(x) > x > \gamma_3$. Similarly, if $x < \gamma_1$, then $\phi(x) < x < \gamma_1$.
If $\gamma_1 + a^{-1/2} < x < \gamma_2 - a^{-1/2}$, then, noting that $\gamma_3 > x$, we have

$$\phi(x) - \gamma_3 = \big[a(x-\gamma_1)(\gamma_2 - x) - 1\big](\gamma_3 - x) > (a(a^{-1/2})^2 - 1)(\gamma_3 - x) \geq 0.$$

Similarly, if $\gamma_2 + a^{-1/2} < x < \gamma_3 - a^{-1/2}$, we obtain $\phi(x) < \gamma_1$. □

Lemma 4.11. *Let $\phi(z) \in \mathbb{R}[z]$ be a cubic polynomial with negative lead coefficient. If $\mathfrak{K}_\infty \cap \mathbb{R}$ consists of more than one point, then ϕ has at least two distinct real periodic points of period two. Moreover, if $\alpha \in \mathbb{R}$ is the smallest such periodic point, then $\phi(\alpha)$ is the largest, and $\mathfrak{K}_\infty \cap \mathbb{R} \subseteq [\alpha, \phi(\alpha)]$.*

Proof. Let $\alpha = \inf(\mathfrak{K}_\infty \cap \mathbb{R})$ and $\beta = \sup(\mathfrak{K}_\infty \cap \mathbb{R})$, so that $\mathfrak{K}_\infty \cap \mathbb{R} \subseteq [\alpha, \beta]$. By hypothesis, $\alpha < \beta$. It suffices to show that $\phi(\alpha) = \beta$ and $\phi(\beta) = \alpha$.

Note that $\phi(\alpha) \in \mathfrak{K}_\infty \cap \mathbb{R}$, and therefore $\phi(\alpha) \leq \beta$. If $\phi(\alpha) < \beta$, then by continuity, there is some $\alpha' < \alpha$ such that $\phi(\alpha') = \beta$, contradicting the minimality of α. Thus, $\phi(\alpha) = \beta$; similarly, $\phi(\beta) = \alpha$. □

5. The search algorithm

We are now ready to describe our algorithm to search for preperiodic points and points of small height for cubic polynomials over $\mathbb{Q}$.

Algorithm 5.1. Given $a \in \mathbb{Q}^\times$ and $b \in \mathbb{Q}$, set $\phi(z) = az^3 + bz + 1$ or $\phi(z) = az^3 + bz$, define $h(\phi)$ as in Definition 4.4, and set $h_+(\phi) = \max\{h(\phi), 1\}$.

1. Let S be the set of all (bad) prime factors p of the numerator of a, denominator of a, and denominator of b. Compute each radius s_p from Lemma 3.3; by Remark 2.3,

$$s_p = \begin{cases} \max\{|b/a|_p^{1/2}, |1/a|_p^{1/2}\} & \text{for } \phi(z) = az^3 + bz, \\ \max\{|b/a|_p^{1/2}, |1/a|_p^{1/3}, |1/a|_p^{1/2}\} & \text{for } \phi(z) = az^3 + bz + 1. \end{cases}$$

Shrink s_p if necessary to be an integer power of p. Let $M = \prod_{p \in S} s_p \in \mathbb{Q}^\times$. Thus, for any preperiodic rational point $x \in \mathbb{Q}$, we have $Mx \in \mathbb{Z}$.

2. If $a > 0$ and ϕ has only one real fixed point, or if $a < 0$ and ϕ has no real two-periodic points, then (by Lemma 4.9 or Lemma 4.11) $\mathfrak{K}_\infty \cap \mathbb{R}$ consists of a single point $\gamma \in \mathbb{R}$, which must be fixed. In that case, check whether γ is rational by

seeing whether $M\gamma$ is an integer; report either the one or zero preperiodic points, and end.

3. Let S' be the set of all $p \in S$ for which $|a|_p^{-1/2} < s_p$. Motivated by Lemma 3.4(4), for each such p consider the (zero, one, two, or three) disks of radius $|a|_p^{-1/2}$ that contain both an element of $\phi^{-1}(0)$ and a $\mathbb{Q}_p$-rational point. If for at least one $p \in S'$ there are no such disks, then report zero preperiodic points, and end.

4. Otherwise, use the Chinese Remainder Theorem to list all rational numbers that lie in the real interval(s) given by Lemma 4.10 or 4.11, are integer multiples of the rational number M from Step 1, and lie in the disks from Step 3 at each $p \in S'$.

5. For each point x in Step 4, compute $\phi^i(x)$ for $i = 0, \ldots, 6$. If any are repeats, record a preperiodic point. Otherwise, compute $h(a(\phi^6(x))^2)/(2 \cdot 3^6 \cdot h_+(\phi))$. If the value is less than .03, record $h(a(\phi^{12}(x))^2)/(2 \cdot 3^{12})$ as $\hat{h}_\phi(x)$, and

$$\mathfrak{h}(x) = \hat{h}_\phi(x)/h_+(\phi) \tag{5-1}$$

as the *scaled height* of x.

Remark 5.2. The definition of $h_+(\phi)$ is designed to avoid dividing by zero when computing $\mathfrak{h}(x)$. In particular, the choice of 1 as a minimum value is arbitrary. Of course, the height $h(\phi)$ already depends on our choice of normal forms in Definition 4.1; moreover, without reference to some kind of canonical structure, Weil heights on varieties are only natural objects up to bounded differences. In other words, $h_+(\phi)$ is no more arbitrary than $h(\phi)$ as a height on the moduli space $\mathcal{M}_3$.

In addition, none of the polynomials we found with points of particularly small scaled height $\mathfrak{h}(x)$ had $h(\phi) \le 1$, even though the change from $h(\phi)$ to $h_+(\phi)$ could only make $\mathfrak{h}(x)$ smaller. Thus, our use of h_+ had no significant effect on the data.

Remark 5.3. Algorithm 5.1 tests only points that, at all places, are in regions where the filled Julia set might be. At nonarchimedean places, that means the region U_1 in Lemma 3.4(4); and at the archimedean place, that means the regions described in Lemma 4.10 or Lemma 4.11. Thus, as mentioned in the discussion following Proposition 3.2, the algorithm is guaranteed to test all preperiodic points, but there is a possibility it may miss a point of small positive height that happens to lie outside the search region at some place. However, such a point must have a nonnegligible positive contribution to its canonical height, coming from the local canonical height at that place.

For example, any point x lying outside the region U_1 at a nonarchimedean place v must satisfy $\phi(x) \notin U_0$. If $p_v \neq 3$, then by Lemma 4.7,

$$U_0 = \bar{D}(0, s_v),$$

outside of which it is easy to show that

$$\hat{\lambda}_{\phi,v}(x) = \lambda_v(x) + \tfrac{1}{2}\log|a|_v.$$

Since $\hat{\lambda}_{\phi,v}(\phi(x)) > 0$ and $\lambda_v(x)$ takes values in $(\log p_v)\mathbb{Z}$, it follows that

$$\hat{\lambda}_{\phi,v}(\phi(x)) \geq \tfrac{1}{2}\log p_v,$$

and therefore

$$\hat{h}_\phi(x) \geq \hat{\lambda}_{\phi,v}(x) \geq \frac{\log p_v}{6} \geq \frac{\log 2}{6} = .1155\ldots.$$

Thus, we are not missing points of height smaller than .11 by restricting to U_1.

Admittedly, at the archimedean place we have no such lower bound, and the possibility exists of missing a point of small height just outside the search region. However, because the denominators of such points (and all their forward iterates!) must divide M, there still cannot be many omitted points of small height unless $h(\phi)$ is very large.

Remark 5.4. The bounds of 6 for preperiodic repeats and .03 for $\mathfrak{h}(x)$, as well as the decision to test $\hat{h}_\phi(x)$ first at 6 iterations and again at 12, were chosen by trial and error. There seem to be many cubic polynomials with points of scaled height smaller than .03, suggesting that our choice of that cutoff is safely large.

Meanwhile, if there happened to be a preperiodic chain of length 7 or longer, our algorithm would not identify the starting point as preperiodic. However, the first point in such a chain would still have shown up in our data as one of extraordinarily small scaled height; but we found no such points in our entire search. That is, *none* of the maps we tested have preperiodic chains of length greater than six.

Finally, by Proposition 4.5 and Theorem 2.7, our preliminary estimate (after six iterations) for $\mathfrak{h}$ is accurate to within $3^{-6}\cdot 1.84 < .0026$, and our sharper estimate (after twelve iterations) is accurate to within $3^{-12}\cdot 1.84 < .0000035$. Thus, the points we test with $\mathfrak{h} < .027$ or $\mathfrak{h} > .033$ cannot be misclassified; and our recorded computations of $\mathfrak{h}$ are accurate to at least five places after the decimal point.

6. Data collected

We ran Algorithm 5.1 on every cubic polynomial $az^3 + bz + 1$ and $az^3 + bz$ for which $a \in \mathbb{Q}^\times$, $b \in \mathbb{Q}$, and both numerators and both denominators are smaller than 300 in absolute value. That means 109,270 choices for a and (because $b = 0$ is allowed) 109,271 choices for b, giving almost 12 billion pairs (a, b). (Not coincidentally, 109,271 is approximately $(12/\pi^2)\cdot 300^2$; see [Silverman 2007, Exercise 3.2(b)].) Of course, in light of Proposition 4.2, we skipped polynomials of the form $\gamma^2 az^3 + bz$ for $\gamma \in \mathbb{Q}$ if we had already tested $az^3 + bz$. That meant only 18,972 choices for a, but the same 109,271 choices for b; as a result, there

	number of form az^3+bz when $h(a), h(b) <$		number of form az^3+bz+1 when $h(a), h(b) <$	
n	$\log 200$	$\log 300$	$\log 200$	$\log 300$
11	10	10	0	0
10	0	0	0	3
9	30	36	20	28
8	0	0	36	52
7	196	318	144	193
6	0	0	257	358
5	524	774	533	751
4	0	0	1,533	2,314
3	132,352	297,826	52,402	115,954
2	0	0	42,447	92,221
1	422,358,932	2,072,790,448	187,391	432,131
0	0	0	2,362,307,079	11,939,398,165
total	422,492,044	2,073,089,412	2,362,591,842	11,940,042,170

Table 1. Number of distinct cubic polynomials of the given forms with $h(a), h(b) < \log 200, \log 300$ and n rational points of scaled height smaller than .03.

were only about 2 billion truly different cubics of the second type. Combining the two families, then, we tested over 14 billion truly different cubic polynomials. We summarize our key observations here; the complete data may be found online at http://www.cs.amherst.edu/~rlb/cubicdata/.

Table 1 lists the number of such polynomials with a prescribed number of points $x \in \mathbb{Q}$ of small height — that is, with $\mathfrak{h} < .03$, where $\mathfrak{h}(x) = \hat{h}_\phi(x)/\max\{h(\phi), 1\}$ is the scaled height of (5-1). It also lists the totals for $h(a), h(b) < \log 200$, for comparison. Of course, every polynomial of the form az^3+bz has an odd number of small height points, by Remark 4.3 and because $x = 0$ is fixed. Meanwhile, there are more polynomials of the form az^3+bz+1 with three small points than with two, because there are several ways to have three preperiodic points (three fixed points, a fixed point with two extra preimages, or a 3-cycle), but essentially only one way to have two: a 2-cycle. After all, a cubic ϕ with two rational fixed points has a third, except in the rare case of multiple roots of $\phi(z) - z$; and if ϕ has a fixed point $\alpha \in \mathbb{Q}$ with a distinct preimage $\beta \in \mathbb{Q}$, then the third preimage is also rational.

According to our data, no cubic polynomial with $h(a), h(b) < \log 300$ has more than 11 rational points of small height. In fact, there are only ten such polynomials

a, b, c	periodic cycles	strictly preperiodic
$-\frac{3}{2},\ \frac{19}{6},\ 0$	$\{\frac{5}{3}, -\frac{5}{3}\}, \{0\}$	$\pm\frac{4}{3}, \pm\frac{2}{3}, \pm\frac{1}{3}, \pm 1$
$\frac{3}{2},\ -\frac{19}{6},\ 0$	$\{\frac{5}{3}\}, \{-\frac{5}{3}\}, \{0\}$	$\pm\frac{4}{3}, \pm\frac{2}{3}, \pm\frac{1}{3}, \pm 1$
$-3,\ \frac{37}{12},\ 0$	$\{\frac{7}{6}, -\frac{7}{6}\}, \{\frac{5}{6}\}, \{-\frac{5}{6}\}, \{0\}$	$\pm\frac{1}{6}, \pm\frac{1}{2}, \pm\frac{2}{3}$
$3,\ -\frac{37}{12},\ 0$	$\{\frac{5}{6}, -\frac{5}{6}\}, \{\frac{7}{6}\}, \{-\frac{7}{6}\}, \{0\}$	$\pm\frac{1}{6}, \pm\frac{1}{2}, \pm\frac{2}{3}$
$-\frac{3}{2},\ \frac{73}{24},\ 0$	$\{\frac{1}{2}, \frac{4}{3}\}, \{-\frac{1}{2}, -\frac{4}{3}\}, \{\frac{7}{6}\}, \{-\frac{7}{6}\}, \{0\}$	$\pm\frac{1}{6}, \pm\frac{3}{2}$
$\frac{3}{2},\ -\frac{73}{24},\ 0$	$\{\frac{1}{2}, -\frac{4}{3}\}, \{-\frac{1}{2}, \frac{4}{3}\}, \{\frac{7}{6}, -\frac{7}{6}\}, \{0\}$	$\pm\frac{1}{6}, \pm\frac{3}{2}$
$-\frac{5}{3},\ \frac{109}{60},\ 0$	$\{\frac{13}{10}, -\frac{13}{10}\}, \{\frac{7}{10}\}, -\{\frac{7}{10}\}, \{0\}$	$\pm\frac{3}{10}, \pm\frac{6}{5}, \pm\frac{1}{2}$
$\frac{5}{3},\ -\frac{109}{60},\ 0$	$\{\frac{7}{10}, -\frac{7}{10}\}, \{\frac{13}{10}\}, \{-\frac{13}{10}\}, \{0\}$	$\pm\frac{3}{10}, \pm\frac{6}{5}, \pm\frac{1}{2}$
$-\frac{6}{5},\ \frac{169}{120},\ 0$	$\{\frac{17}{12}, -\frac{17}{12}\}, \{\frac{7}{12}\}, \{-\frac{7}{12}\}, \{0\}$	$\pm\frac{13}{12}, \pm\frac{2}{3}, \pm\frac{5}{4}$
$\frac{6}{5},\ -\frac{169}{120},\ 0$	$\{\frac{7}{12}, -\frac{7}{12}\}, \{\frac{17}{12}\}, \{-\frac{17}{12}\}, \{0\}$	$\pm\frac{13}{12}, \pm\frac{2}{3}, \pm\frac{5}{4}$
$\frac{1}{240}, -\frac{151}{60},\ 1$	$\{-10, 22\}, \{12, -22\}, \{18, -20\}$	$10, -18, \pm 28$
$-\frac{1}{240}, \frac{151}{60},\ 1$	—	—
$-\frac{169}{240}, \frac{259}{60},\ 1$	$\{-2\}, \{-\frac{4}{13}\}, \{\frac{30}{13}\}$	$\frac{4}{13}, -\frac{10}{13}, -\frac{30}{13}, \pm\frac{34}{13}, \frac{36}{13}$

Table 2. Cubic polynomials az^3+bz+c with ten or more points of small height. The nonpreperiodic points of small positive height pertaining to the last two rows are as follows. Penultimate row: $-12, 20$ with $\mathfrak{h} = .00244$, then $10, 18, -22, -28$ with $\mathfrak{h} = .00733$, then $-10, -18, 22, 28$ with $\mathfrak{h} = .02198$. Last row: $-\frac{14}{13}$ with $\mathfrak{h} = .02947$. Other rows have no such points.

with 11 small points; these are shown in the upper portion of Table 2. All ten have only preperiodic points as points of small height; all have $h(a), h(b) < \log 200$; and all are in the $az^3 + bz$ family. (Five are negatives of the other five, and the negative of any preperiodic point is also preperiodic, as discussed in Remark 4.3.)

Table 2 also lists the only three polynomials in our search with exactly ten points of small height. A complete list, ordered by $h(\phi)$, of those with exactly nine points of small height can be found in Table 3 ($c = 0$) and Table 4 ($c = 1$).

Remark 6.1. Most of the points sharing the same canonical height in Tables 3 and 4 do so simply because one or two iterates later, they coincide. For example, consider the fourth map in Table 4, namely $\phi(z) = \frac{3}{8}z^3 - \frac{49}{24}z + 1$. The three points $0, \pm\frac{7}{3}$ all satisfy $\phi(x) = 1$, and hence all three have the same canonical height.

a, b	periodic cycles	strictly preperiodic	small height points	> 0 $\hat{\mathfrak{h}}$
$\frac{3}{2}, -\frac{13}{6}$	$\{0\}, \{\frac{2}{3}, -1\}, \{-\frac{2}{3}, 1\}$	$\pm\frac{1}{3}, \pm\frac{4}{3}$	—	
$3, -\frac{13}{12}$	$\{0\}, \{\frac{1}{6}, -\frac{1}{6}\}, \{\frac{5}{6}\}, \{-\frac{5}{6}\}$	$\pm\frac{1}{2}, \pm\frac{2}{3}$	—	
$\frac{5}{3}, -\frac{49}{15}$	$\{0\}, \{\frac{8}{5}\}, \{-\frac{8}{5}\}$	$\pm 1, \pm\frac{3}{5}, \pm\frac{7}{5}$	—	
$\frac{3}{2}, -\frac{49}{24}$	$\{0\}, \{\frac{5}{6}, -\frac{5}{6}\}$	$\pm\frac{1}{2}, \pm\frac{4}{3}, \pm\frac{7}{6}$	—	
$\frac{5}{2}, -\frac{49}{40}$	$\{0\}, \{\frac{3}{10}, -\frac{3}{10}\}$	$\pm\frac{1}{2}, \pm\frac{4}{5}, \pm\frac{7}{10}$	—	
$\frac{6}{5}, -\frac{61}{30}$	$\{0\}, \{1, -\frac{5}{6}\}, \{-1, \frac{5}{6}\}$	$\pm\frac{2}{3}, \pm\frac{3}{2}$	—	
$\frac{6}{5}, -\frac{79}{30}$	$\{0\}, \{\frac{7}{6}, -\frac{7}{6}\}$	$\pm\frac{1}{2}, \pm\frac{5}{3}$	$\pm\frac{1}{3}$	.02046
$\frac{6}{5}, -\frac{91}{30}$	$\{0\}, \{\frac{11}{6}\}, \{-\frac{11}{6}\}$	$\pm 1, \pm\frac{5}{6}$	$\pm\frac{2}{3}$	.01982
$\frac{6}{5}, -\frac{139}{30}$	$\{0\}, \{\frac{13}{6}\}, \{-\frac{13}{6}\}$	$\pm\frac{1}{2}, \pm\frac{5}{3}$	± 2	.02525
$\frac{7}{6}, -\frac{163}{42}$	$\{0\}, \{\frac{11}{7}, -\frac{11}{7}\}$	$\pm 2, \pm\frac{3}{7}, \pm\frac{4}{7}$	—	
$\frac{7}{2}, -\frac{169}{56}$	$\{0\}, \{\frac{15}{14}\}, \{-\frac{15}{14}\}$	$\pm\frac{1}{2}, \pm\frac{4}{7}, \pm\frac{13}{14}$	—	
$\frac{5}{3}, -\frac{169}{60}$	$\{0\}, \{\frac{1}{2}, -\frac{6}{5}\}, \{-\frac{1}{2}, \frac{6}{5}\}$	$\pm\frac{13}{10}$	$\pm\frac{3}{10}$	.02744
$\frac{15}{7}, -\frac{169}{105}$	$\{0\}, \{\frac{8}{15}, -\frac{8}{15}\}$	$\pm 1, \pm\frac{7}{15}, \pm\frac{13}{15}$	—	
$\frac{5}{3}, -\frac{181}{60}$	$\{0\}, \{\frac{11}{10}, -\frac{11}{10}\}$	$\pm\frac{3}{2}, \pm\frac{2}{5}, \pm\frac{9}{10}$	—	
$\frac{7}{3}, -\frac{193}{84}$	$\{0\}, \{\frac{1}{2}, -\frac{6}{7}\}, \{-\frac{1}{2}, \frac{6}{7}\}$	$\pm\frac{8}{7}, \pm\frac{9}{14}$	—	
$\frac{5}{3}, -\frac{229}{60}$	$\{0\}, \{\frac{13}{10}, -\frac{13}{10}\}, \{\frac{17}{10}\}, \{-\frac{17}{10}\}$	$\pm\frac{6}{5}, \pm\frac{1}{2}$	—	
$\frac{5}{3}, -\frac{241}{60}$	$\{0\}, \{\frac{3}{2}, -\frac{2}{5}\}, \{-\frac{3}{2}, \frac{2}{5}\}$	$\pm\frac{8}{5}, \pm\frac{1}{10}$	—	
$\frac{6}{5}, -\frac{289}{120}$	$\{0\}, \{\frac{2}{3}, -\frac{5}{4}\}, \{-\frac{2}{3}, \frac{5}{4}\}, \{\frac{13}{12}, -\frac{13}{12}\}$	$\pm\frac{17}{12}$	—	

Table 3. Cubic polynomials az^3+bz with $a > 0$ and nine points of small height. To save space, only polynomials az^3+bz with $a > 0$ are listed; to obtain those with $a < 0$, simply replace each pair (a, b) by $(-a, -b)$ and adjust the cycle structure of the periodic points according to Remark 4.3.

Meanwhile, $\phi(-\frac{1}{3}) = \frac{5}{3} \neq 1$, but $\phi(1) = \phi(\frac{5}{3}) = -\frac{2}{3}$, and hence $-\frac{1}{3}$ also has the same common canonical height.

The map $\phi(z) = -\frac{27}{80}z^3 + \frac{151}{60}z + 1$, near the bottom of Table 4 (page 59), is an exception to this trend. The points -2, $-\frac{10}{9}$, and $\frac{28}{9}$ all satisfy $\phi(x) = -\frac{4}{3}$, but

a, b	periodic cycles	strictly preperiodic	small height >0 points	$\mathfrak{h}$
$\frac{1}{6}, -\frac{13}{6}$	$\{3, -1\}$	$0, 1, \pm 2, -3, \pm 4$	—	
$-\frac{1}{6}, \frac{13}{6}$	$\{3\}, \{-1\}, \{-2\}$	$0, 1, 2, -3, \pm 4$	—	
$\frac{3}{4}, -\frac{49}{12}$	$\{1, -\frac{7}{3}\}$	$0, \pm\frac{1}{3}, \frac{4}{3}, \frac{5}{3}, \frac{7}{3}, -\frac{8}{3}$	—	
$\frac{3}{8}, -\frac{49}{24}$	$\{-3\}, \{\frac{1}{3}\}, \{\frac{8}{3}\}$	$-1, -\frac{5}{3}$	$0, -\frac{1}{3}, \pm\frac{7}{3}$	.02309
$\frac{25}{24}, -\frac{49}{24}$	$\{0, 1\}$	$\pm\frac{1}{5}, \frac{3}{5}, \pm\frac{7}{5}, -\frac{8}{5}, -\frac{9}{5}$	—	
$\frac{5}{12}, -\frac{49}{60}$	$\{\frac{3}{5}\}$	$0, \pm 1, -\frac{3}{5}, \pm\frac{7}{5}, \pm\frac{8}{5}$	—	
$-\frac{5}{12}, \frac{49}{60}$	$\{1, \frac{7}{5}\}$	$0, -1, -\frac{7}{5}, \pm\frac{3}{5}, \pm\frac{8}{5}$	—	
$-\frac{49}{48}, \frac{19}{12}$	$\{\frac{12}{7}, -\frac{10}{7}\}, \{\frac{2}{7}, \frac{10}{7}\}$	$-\frac{2}{7}, \pm\frac{4}{7}, \pm\frac{6}{7}$	—	
$\frac{2}{15}, -\frac{91}{30}$	$\{\frac{5}{2}, -\frac{9}{2}\}$	$3, \pm 5, \pm\frac{1}{2}, \frac{9}{2}, -\frac{11}{2}$	—	
$-\frac{1}{30}, \frac{91}{30}$	$\{4, 11, -10\}$	$0, 1, 9, -5, -6$	-4	.01983
$\frac{1}{48}, -\frac{31}{12}$	$\{-4, 10\}, \{-10, 6\}$	$\pm 2, -6, \pm 12$	—	
$-\frac{1}{48}, \frac{31}{12}$	—	—	4 $2, 10, -12$ ± 6 $-2, -10, 12$	.00039 .00118 .00355 .01065
$\frac{49}{48}, -\frac{31}{12}$	$\{-2\}, \{\frac{2}{7}\}, \{\frac{12}{7}\}$	$-\frac{2}{7}, \frac{4}{7}, \pm\frac{10}{7}, -\frac{12}{7}$	$\frac{6}{7}$	.02587
$\frac{3}{16}, -\frac{43}{12}$	$\{-4, \frac{10}{3}\}$	$-\frac{2}{3}, \frac{14}{3}$	2 $\frac{4}{3}$ $4, \frac{2}{3}, -\frac{14}{3}$	.00677 .01653 .02030
$-\frac{3}{16}, \frac{43}{12}$	$\{-4, -\frac{4}{3}, -\frac{10}{3}\}$	$\frac{14}{3}, -\frac{2}{3}$	-2 $4, \frac{2}{3}, -\frac{14}{3}$	.00488 .01463
$\frac{3}{40}, -\frac{241}{120}$	$\{\frac{1}{3}\}$	$\pm 3, 5, -\frac{16}{3}, -\frac{19}{3}$	$-\frac{1}{3}, -5, \frac{16}{3}$	.02182
$-\frac{3}{40}, \frac{241}{120}$	$\{-3\}$	—	$\left.\begin{matrix} 3, -5, -\frac{1}{3} \\ \frac{16}{3}, \frac{19}{3} \end{matrix}\right\}$ $5, \frac{1}{3}, -\frac{16}{3}$	.00933 .02800
$\frac{27}{80}, -\frac{91}{60}$	$\{\frac{4}{3}, -\frac{2}{9}\}$	$-2, -\frac{4}{3}, \pm\frac{10}{9}, \frac{20}{9}, \pm\frac{22}{9}$	—	

Table 4. Cubic polynomials $az^3 + bz + 1$ with nine points of small height (continued on next page).

a, b	periodic cycles	strictly preperiodic	small height > 0 points	$\mathfrak{h}$
$-\frac{27}{80}, \frac{91}{60}$	—	—	$2, \frac{2}{9}, -\frac{20}{9}$ $\pm\frac{4}{3}, \pm\frac{10}{9}, \pm\frac{22}{9}$	.00505 .01516
$\frac{121}{80}, -\frac{91}{20}$	$\{-2\}, \{\frac{2}{11}\}, \{\frac{20}{11}\}$	$-\frac{2}{11}, \frac{10}{11}, \frac{12}{11}, \pm\frac{18}{11}, -\frac{20}{11}$	—	
$\frac{1}{240}, -\frac{91}{60}$	$\{-10, 12\}$	$10, -12, \pm 22$	$2, 18, -20$	.01806
$-\frac{1}{240}, \frac{91}{60}$	$\{-2\}, \{-10\}, \{12\}$	$10, -12, -18, 20, \pm 22$	—	
$\frac{169}{96}, -\frac{133}{24}$	$\{-2\}, \{\frac{2}{13}\}, \{\frac{24}{13}\}$	$-\frac{2}{13}, \frac{8}{13}, \frac{18}{13}, \pm\frac{22}{13}, -\frac{24}{13}$	—	
$-\frac{289}{240}, \frac{139}{60}$	$\{\frac{4}{17}, \frac{26}{17}\}, \{-\frac{26}{17}, \frac{30}{17}\}$	$\pm\frac{6}{17}, \pm\frac{20}{17}$	$-\frac{14}{17}$	.02485
$-\frac{27}{80}, \frac{151}{60}$	$\{\frac{10}{3}, -\frac{28}{9}\}$	$2, \frac{10}{9}$	$-2, -\frac{10}{9}$ $\frac{28}{9}, -\frac{22}{9}$ $\frac{22}{9}$	.01396 .01418
$\frac{3}{112}, -\frac{247}{84}$	$\{12\}$	$2, -\frac{14}{3}, -\frac{22}{3}, \frac{28}{3}, -\frac{34}{3}$	$-2, -\frac{28}{3}, \frac{34}{3}$	.02313
$-\frac{3}{112}, \frac{247}{84}$	—	—	$-2, -12, \frac{14}{3}$ $\frac{22}{3}, -\frac{28}{3}, \frac{34}{3}$ $2, \frac{28}{3}, -\frac{34}{3}$	.01568 .01995
$\frac{3}{80}, -\frac{259}{60}$	$\{-12\}$	$\frac{10}{3}, \frac{26}{3}$	$-\frac{4}{3}, -10, \frac{34}{3}$ $\frac{4}{3}, 10, -\frac{34}{3}$	.02012 .02974

Table 4 (continued). Cubic polynomials $az^3 + bz + 1$ with nine points of small height.

all iterates of $-\frac{22}{9}$ appear to be distinct from those of -2. Nonetheless, all four points share the same canonical height $\hat{h}_\phi(-\frac{22}{9}) = \hat{h}_\phi(-2) = \frac{1}{18}\log 5 \approx .08941$. (The scaled height .01396 is of course .08941 divided by $h(\phi) = \log 604$.) We can compute this explicit value as follows. The bad primes are $v = 2, 3, 5, \infty$. In $\mathbb{R}$, the iterates of all four points approach the fixed point at -1.639. At $v = 3$, ϕ maps the set $\{x \in \mathbb{Q}_3 : |x|_3 \le 9\}$ into itself, since $9\phi(z/9) = \frac{1}{3}(z^3 - z) - \frac{27}{80}z^3 + \frac{57}{20}z + 9$ maps 3-adic integers to 3-adic integers. At $v = 2$, one can show that ϕ maps $\bar{D}(4, \frac{1}{16})$ into $\bar{D}(2, \frac{1}{4})$, $\bar{D}(2, \frac{1}{8})$ into $\bar{D}(-2, \frac{1}{8})$, $\bar{D}(-2, \frac{1}{16})$ into $\bar{D}(4, \frac{1}{16})$, and $\bar{D}(6, \frac{1}{16})$ into $\bar{D}(4, \frac{1}{16})$; hence the orbit any point $x \in \mathbb{Q}_2$ in these disks stays in the same disks. Thus, $\hat{\lambda}_{\phi,\infty}(x) = \hat{\lambda}_{\phi,3}(x) = \hat{\lambda}_{\phi,2}(x) = 0$ for all four points x; by Propositions 1.2 and 1.4, then, $\hat{h}_\phi(x) = \hat{\lambda}_{\phi,5}(x)$. Finally, all four points satisfy $|\phi^3(x)|_5 = 5$, and therefore $|\phi^n(x)|_5 = 5^{e_n}$ for all $n \geq 3$, where

$$e_n = 1 + 3 + \cdots + 3^{n-3} = \frac{3^{n-2} - 1}{2}.$$

Thus, $\hat{\lambda}_{\phi,5}(x) = \lim_{n\to\infty}(e_n/3^n)\log 5 = \frac{1}{18}\log 5$, as claimed. Incidentally, this same argument almost applies to the fifth point $x = \frac{22}{9}$ as well, but $\phi^7(\frac{22}{9}) \approx 36.19$. As a result, $\hat{\lambda}_{\phi,\infty}(\frac{22}{9}) \approx .0014$ is positive; dividing by $\log 604$ gives the extra contribution of .00022 to the scaled height.

A similar phenomenon occurs for the map $\phi(z) = -\frac{1}{240}z^3 + \frac{259}{60}z + 1$, listed near the bottom of Table 2. For that map, -12 and 20 have the same small height but apparently disjoint orbits. The point 20 maps to 18, and all three of 10, 18, and -28 map to 22, which then maps to 12. Meanwhile, -12 maps to -22, which maps to -10; and all three of -10, -18, and 28 map to -20.

As mentioned in Remark 5.4, no preperiodic point in our data took more than six iterations to produce a repeated value. In fact, all but one function required only five. The one exception is $\phi(z) = \frac{1}{12}z^3 - \frac{25}{12}z + 1$, for which the preperiodic point 0 lands on the 5-periodic point 1 after one iteration. (There are a total of 8 small height points for ϕ, because -5 also maps to 1, and because -4 has scaled height .01595 . . . .) This map was also the only cubic polynomial in our search with a rational 5-periodic point; all other periods were at most 4. Table 5 lists all those cubic polynomials in our search for which some rational preperiodic point required 5 or more iterations to reach a repeat; note that all are of the form $az^3 + bz + 1$.

Our data supports Conjecture 1 for cubic polynomials inasmuch as the number of rational preperiodic points does not grow as $h(\phi)$ increases. For example, even though Table 2 shows a number of maps $az^3 + bz$ with eleven preperiodic points, it is important to note that the first such map had height as small as $h(\phi) = \log 19$. Similarly, every preperiodic structure appearing anywhere in Tables 2, 3, and 4 appeared already for some map of relatively small height. That is, the data suggests that all the phenomena that can occur have already occurred among the small height maps.

In the same way, the data also supports Conjecture 2 for cubic polynomials. Table 6 lists the only nine points of scaled height smaller than .0007 in our entire search. (There were only twenty points with scaled height smaller than .001; three of the extra eleven are iterates of the first three points listed in Table 6.) Once again, even though there are two maps of fairly large height ($\log 289 \approx 5.67$ and $\log(27 \cdot 12) \approx 5.78$) with a point of small scaled height, there was already a map of substantially smaller height ($\log 97 \approx 3.37$) with an even smaller point.

Moreover, the intuition (mentioned in the introduction) that the scaled height measures the number of iterates required to start the "explosion" is on clear display in Table 6. For these points, it takes seven applications of ϕ to get to an iterate with

a, b, c	$x, \phi(x), \phi^2(x), \ldots$	other preperiodic	small height > 0 point	$\mathfrak{h}$
$\frac{1}{12}, -\frac{25}{12}, 1$	$0, \{1, -1, 3, -3, 5\}$	-5	-4	.01595
$-\frac{3}{2}, \frac{7}{6}, 1$	$0, 1, \frac{2}{3}, \{\frac{4}{3}, -1\}$	$\pm\frac{1}{3}, -\frac{2}{3}$	—	
$\frac{1}{6}, -\frac{13}{6}, 1$	$2, -2, 4, \{3, -1\}$	$0, 1, -3, -4$	—	
$\frac{2}{3}, -\frac{13}{6}, 1$	$\frac{3}{2}, 0, 1, -\frac{1}{2}, \{2\}$	$-2, \frac{1}{2}, -\frac{3}{2}$	—	
$-\frac{4}{3}, \frac{13}{12}, 1$	$0, 1, \{\frac{3}{4}, \frac{5}{4}, -\frac{1}{4}\}$	$-1, \frac{1}{4}, -\frac{3}{4}$	—	
$\frac{3}{4}, -\frac{25}{12}, 1$	$\frac{4}{3}, 0, \{1, -\frac{1}{3}, \frac{5}{3}\}$	$\frac{1}{3}, -\frac{5}{3}$	—	
$-\frac{1}{3}, \frac{37}{12}, 1$	$\frac{1}{2}, \frac{5}{2}, \frac{7}{2}, \{-\frac{5}{2}, -\frac{3}{2}\}$	$-2, -\frac{1}{2}$	—	
$-\frac{4}{3}, \frac{37}{12}, 1$	$-\frac{5}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{7}{4}, \{-\frac{3}{4}\}$	-1	—	
$\frac{3}{4}, -\frac{49}{12}, 1$	$\frac{1}{3}, -\frac{1}{3}, \frac{7}{3}, \{1, -\frac{7}{3}\}$	$0, \frac{4}{3}, \frac{5}{3}, -\frac{8}{3}$	—	
$\frac{6}{5}, -\frac{61}{30}, 1$	$-\frac{3}{2}, \{0, 1, \frac{1}{6}, \frac{2}{3}\}$	$\frac{5}{6}$	—	
$-\frac{32}{3}, \frac{37}{6}, 1$	$-\frac{5}{8}, -\frac{1}{4}, -\frac{3}{8}, \{-\frac{3}{4}, \frac{7}{8}\}$	$-\frac{1}{2}$	—	
$\frac{1}{48}, -\frac{19}{12}, 1$	$2, -2, 4, \{-4, 6\}$	$-6, \pm 10$	—	
$-\frac{2}{15}, \frac{79}{30}, 1$	$-4, -1, -\frac{3}{2}, \{-\frac{5}{2}, -\frac{7}{2}\}$	5	$\begin{cases} 0 \\ 1 \end{cases}$	.00688 .02065
$-\frac{1}{30}, \frac{91}{30}, 1$	$0, 1, \{4, 11, -10\}$	$-5, -6, 9$	-4	.01983
$\frac{8}{15}, -\frac{121}{30}, 1$	$\frac{1}{4}, 0, \{1, -\frac{5}{2}, \frac{11}{4}\}$	$-\frac{11}{4}$	—	
$-\frac{49}{48}, \frac{31}{12}, 1$	$-\frac{6}{7}, -\frac{4}{7}, -\frac{2}{7}, \{\frac{2}{7}, \frac{12}{7}\}$	$\pm\frac{10}{7}, -\frac{12}{7}$	—	
$\frac{5}{48}, -\frac{211}{60}, 1$	$\frac{22}{5}, -\frac{28}{5}, \{\frac{12}{5}, -6, -\frac{2}{5}\}$	$6, \frac{28}{5}, \frac{2}{5}$	—	

Table 5. Cubic polynomials $az^3 + bz + c$ having a rational preperiodic chain of length ≥ 5.

noticeably larger numerator or denominator than its predecessors. To get a point of smaller scaled height than the record of .00025 in Table 6, then, it seems one would need a point and map with *eight* iterations required to start the explosion.

Also of note is that, just as in Table 5, all the maps in Table 6 are of the form $az^3 + bz + 1$. In fact, the smallest scaled height for a map $az^3 + bz$ occurs for $\pm\frac{5}{3}z^3 \mp \frac{77}{30}z$, at $x = \pm\frac{4}{5}$. (Once again, see Remark 4.3 to explain the four-way tie.) The scaled height is .00591, more than twenty times as large as the current record for $az^3 + bz + 1$; indeed, it takes a mere four iterations to land on $43/40$, at which point the numerator and denominator both start to explode.

This phenomenon supports the heuristic behind Conjectures 1 and 2, that it is

a, b, c	$h(\phi)$	$x, \phi(x), \phi^2(x), \ldots$	$\mathfrak{h}(x)$
$-\frac{25}{24}, \frac{97}{24}, 1$	3.3672	$-\frac{7}{5}, -\frac{9}{5}, -\frac{1}{5}, \frac{1}{5}, \frac{9}{5}, \frac{11}{5}, -\frac{6}{5}, -\frac{41}{20}, \frac{4323}{2560}, \ldots$	.00025
$\frac{8}{15}, -\frac{289}{120}, 1$	5.6664	$\frac{5}{8}, -\frac{3}{8}, \frac{15}{8}, 0, 1, -\frac{7}{8}, \frac{11}{4}, \frac{175}{32}, \frac{307441}{4096}, \ldots$	.00030
$-27, \frac{85}{12}, 1$	5.7807	$-\frac{2}{9}, -\frac{5}{18}, -\frac{7}{18}, -\frac{1}{6}, -\frac{1}{18}, \frac{11}{18}, -\frac{5}{6}, \frac{193}{18}, -\frac{597703}{18}, \ldots$	.00032
$-\frac{1}{48}, \frac{31}{12}, 1$	4.8202	$4, 10, 6, 12, -4, -8, -9, -\frac{113}{16}, -\frac{649189}{65536}, \ldots$	.00039
$-\frac{3}{4}, \frac{25}{12}, 1$	3.2188	$-1, -\frac{1}{3}, \frac{1}{3}, \frac{5}{3}, 1, \frac{7}{3}, -\frac{11}{3}, \frac{91}{3}, -\frac{62605}{3}, \ldots$	.00046
$\frac{21}{128}, -\frac{295}{168}, 1$	8.4596	$-4, -\frac{52}{21}, \frac{20}{7}, -\frac{4}{21}, \frac{4}{3}, -\frac{20}{21}, \frac{124}{49}, -\frac{39572}{50421}, \ldots$	.00047
$-\frac{243}{224}, \frac{85}{168}, 1$	6.5917	$-\frac{2}{27}, \frac{26}{27}, \frac{14}{27}, \frac{10}{9}, \frac{2}{27}, \frac{28}{27}, \frac{17}{54}, \frac{2593}{2304}, \frac{2336653975}{101468602368}, \ldots$	.00057
$\frac{4}{21}, -\frac{205}{84}, 1$	5.3230	$\frac{3}{4}, -\frac{3}{4}, \frac{11}{4}, -\frac{7}{4}, \frac{17}{4}, \frac{21}{4}, \frac{63}{4}, \frac{2827}{4}, \frac{1882717007}{28}, \ldots$	.00058
$\frac{15}{8}, -\frac{289}{120}, 1$	5.6664	$\frac{1}{5}, \frac{8}{15}, 0, 1, \frac{7}{15}, \frac{1}{15}, \frac{21}{25}, \frac{276}{3125}, \frac{9626315307}{12207031250}, \ldots$	.00063

Table 6. Cubic polynomials $az^3 + bz + c$ with rational points of scaled height less than .0007.

hard to have a lot of points of small height, as follows. If x were a small height point for $az^3 + bz$, then $-x$ would have the same small height; their iterates would also have (not quite as) small heights, too. Together with the fixed point at 0, then, there would be more small height points than the heuristic would say are allowed. This idea is further supported by Tables 2, 3, and 4: while it is possible to have eleven preperiodic points or ten points of small height, or even some of each, it does not seem possible to have more than eleven total such points. Thus, there seems to be an upper bound for the total number of points of small height, as predicted by Conjectures 1 and 2.

Acknowledgements

The results of this paper came from an REU in Summer 2007 at Amherst College, supported by the NSF and Amherst College. Authors Dickman, Joseph, and Rubin, who were the REU participants, as well as Benedetto, who supervised the project, gratefully acknowledge the support of NSF grant DMS-0600878. The remaining authors gratefully acknowledge the support of Amherst College: the Schupf Scholar Program for Krause, and Dean of Faculty student funds for Zhou. The authors also thank Amherst College, and especially Scott Kaplan, for the use of the College's high-performance computing cluster, on which we ran our computations. Finally, we thank Patrick Ingram, Bjorn Poonen, Joseph Silverman, Franco Vivaldi, and the referee for their helpful comments on the exposition.

References

[Baker 2006] M. Baker, "A lower bound for average values of dynamical Green's functions", *Math. Res. Lett.* **13**:2-3 (2006), 245–257. MR 2007c:11078

[Benedetto 2007] R. L. Benedetto, "Preperiodic points of polynomials over global fields", *J. Reine Angew. Math.* **608** (2007), 123–153. MR 2008j:11071

[Call and Goldstine 1997] G. S. Call and S. W. Goldstine, "Canonical heights on projective space", *J. Number Theory* **63**:2 (1997), 211–243. MR 98c:11060 Zbl 0895.14006

[Call and Silverman 1993] G. S. Call and J. H. Silverman, "Canonical heights on varieties with morphisms", *Compositio Math.* **89**:2 (1993), 163–205. MR 95d:11077 Zbl 0826.14015

[Flynn et al. 1997] E. V. Flynn, B. Poonen, and E. F. Schaefer, "Cycles of quadratic polynomials and rational points on a genus-2 curve", *Duke Math. J.* **90**:3 (1997), 435–463. MR 98j:11048 Zbl 0958.11024

[Gillette 2004] A. Gillette, *Searching for small canonical heights: finding a minumum speed limit on the road to infinity*, undergraduate thesis, Amherst College, 2004.

[Gouvêa 1997] F. Q. Gouvêa, *p-adic numbers: an introduction*, Second ed., Springer, Berlin, 1997. MR 98h:11155 Zbl 0874.11002

[Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry: an introduction*, Graduate Texts in Math. **201**, Springer, New York, 2000. MR 2001e:11058 Zbl 0948.11023

[Ingram $\geq$ 2009] P. Ingram, "Lower bounds on the canonical height associated to the morphism $\phi(z) = z^d + c$", *Monatsh. Math.*. To appear.

[Koblitz 1984] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions*, 2nd ed., Graduate Texts in Math. **58**, Springer, New York, 1984. MR 86c:11086

[Lang 1983] S. Lang, *Fundamentals of Diophantine geometry*, Springer-Verlag, New York, 1983. MR 85j:11005 Zbl 0528.14013

[Manes 2007] M. Manes, *Arithmetic dynamics of rational maps*, Ph.D. thesis, Brown University, 2007.

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186 (1978). MR 80c:14015 Zbl 0394.14008

[Merel 1996] L. Merel, "Bornes pour la torsion des courbes elliptiques sur les corps de nombres", *Invent. Math.* **124**:1-3 (1996), 437–449. MR 96i:11057 Zbl 0936.11037

[Morton 1992] P. Morton, "Arithmetic properties of periodic points of quadratic maps", *Acta Arith.* **62**:4 (1992), 343–372. MR 93k:12004 Zbl 0767.11016

[Morton 1998] P. Morton, "Arithmetic properties of periodic points of quadratic maps, II", *Acta Arith.* **87**:2 (1998), 89–102. MR 2000c:11103 Zbl 1029.12002

[Morton and Silverman 1994] P. Morton and J. H. Silverman, "Rational periodic points of rational functions", *Internat. Math. Res. Notices* **1994**:2 (1994), 97–110. MR 95b:11066 Zbl 0819.11045

[Morton and Silverman 1995] P. Morton and J. H. Silverman, "Periodic points, multiplicities, and dynamical units", *J. Reine Angew. Math.* **461** (1995), 81–122. MR 96b:11090 Zbl 0813.11059

[Narkiewicz 1989] W. Narkiewicz, "Polynomial cycles in algebraic number fields", *Colloq. Math.* **58**:1 (1989), 151–155. MR 90k:11135 Zbl 0703.12002

[Northcott 1950] D. G. Northcott, "Periodic points on an algebraic variety", *Ann. of Math.* (2) **51** (1950), 167–177. MR 11,615c Zbl 0036.30102

[Pezda 1994] T. Pezda, "Polynomial cycles in certain local domains", *Acta Arith.* **66**:1 (1994), 11–22. MR 95a:11102 Zbl 0803.11063

[Poonen 1998] B. Poonen, "The classification of rational preperiodic points of quadratic polynomials over $\mathbb{Q}$: a refined conjecture", *Math. Z.* **228**:1 (1998), 11–29. MR 99j:11076 Zbl 0902.11025

[Silverman 2007] J. H. Silverman, *The arithmetic of dynamical systems*, Graduate Texts in Math. **241**, Springer, New York, 2007. MR 2008c:11002 Zbl 1130.37001

[Zieve 1996] M. Zieve, *Cycles of polynomial mappings*, Ph.D. thesis, UC Berkeley, 1996.

Received: 2008-09-25 Revised: 2008-11-25 Accepted: 2008-11-26

rlb@cs.amherst.edu
Amherst College, Department of Math and Computer Science, 31 Quadrangle Drive, Amherst, MA 01002, United States
http://www.cs.amherst.edu/~rlb

bdickman08@amherst.edu
Amherst College, Department of Math and Computer Science, 31 Quadrangle Drive, Amherst, MA 01002, United States

sjoseph08@amherst.edu
Amherst College, Department of Math and Computer Science, 31 Quadrangle Drive, Amherst, MA 01002, United States

bkrause10@amherst.edu
Amherst College, Department of Math and Computer Science, 31 Quadrangle Drive, Amherst, MA 01002, United States

drubin9@jhu.edu
Johns Hopkins University, Department of Mathematics, 3400 N. Charles St, Baltimore, MD 21218, United States

xzhou09@amherst.edu
Amherst College, Department of Math and Computer Science, 31 Quadrangle Drive, Amherst, MA 01002, United States

INVOLVE 2:1(2009)

Divisor concepts for mosaics of integers

Kristen Bildhauser, Jared Erickson, Cara Tacoma and Rick Gillman

(Communicated by Andrew Granville)

The mosaic of the integer n is the array of prime numbers resulting from iterating the Fundamental Theorem of Arithmetic on n and on any resulting composite exponents. In this paper, we generalize several number theoretic functions to the mosaic of n, first based on the primes of the mosaic, second by examining several possible definitions of a divisor in terms of mosaics. Having done so, we examine properties of these functions.

1. Introduction

Mosaics. Mullin, in a series of papers [1964, 1965, 1967a, 1967b], introduced the number theoretic concept of the mosaic of n and explored several ideas related to it.

Definition 1.1. The *mosaic of the integer n* is the array of prime numbers resulting from iterating the Fundamental Theorem of Arithmetic (FTA) on n and on any resulting composite exponents.

The following example illustrates this definition.

$$\begin{aligned} n &= 1{,}024{,}000{,}000 && \leftarrow\text{-- use the FTA to find the prime factorization of } n \\ &= 2^{16}\cdot 5^{6} && \leftarrow\text{-- apply FTA to composite exponents 16 and 6} \\ &= 2^{2^4}\cdot 5^{2\cdot 3} && \leftarrow\text{-- apply FTA again to composite number 4} \\ &= 2^{2^{2^2}}\cdot 5^{2\cdot 3}. && \leftarrow\text{-- the mosaic of the integer; only primes remain} \end{aligned}$$

Mullin introduced several functions on the mosaic, the first of which was $\psi(n)$, the product of all of the primes in the mosaic of n. As an example, we have

$$\psi\left(2^{17^3}\cdot 3^{5^5}\right) = 2\cdot 17\cdot 3\cdot 3\cdot 5\cdot 5 = 7650.$$

MSC2000: 11A99, 11A25, 11A05.

Keywords: number theory, mosaic, factorization.

This material is based upon work supported by the National Science Foundation under Grant No. 033870. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was accomplished while Erickson was affiliated with Valparaiso University.

In number theory, a function is multiplicative if and only if $f(mn) = f(m)f(n)$ whenever m and n are relatively prime. Mullin extended this concept to mosaics by saying that f is *generalized multiplicative* if and only if $f(mn) = f(m)f(n)$ whenever the mosaics of m and n have no primes in common. He showed that $\psi(n)$ is generalized multiplicative and that any multiplicative function is also generalized multiplicative. He also generalized the Möbius function $\mu(n)$ to the *generalized Möbius function*:

$$\mu^*(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if the mosaic of } n > 1 \text{ has any prime number repeated,} \\ (-1)^k & \text{if the mosaic of } n > 1 \text{ has no prime repeated, where } k \text{ is} \\ & \quad \text{the number of \textit{distinct} primes in the mosaic of } n. \end{cases}$$

Similarly, Mullin generalized the concept of additivity: a function f is *generalized additive* if and only if $f(mn) = f(m) + f(n)$ whenever the mosaics of m and n have no primes in common. He defined $\psi^*(n)$ as the sum of the primes in the mosaic of n and showed that this function was generalized additive. As an example,

$$\psi^*\big(5^{2^3\cdot 7}\cdot 11^{13^{19}}\big) = 5+2+3+7+11+13+19 = \psi^*\big(5^{2^3\cdot 7}\big) + \psi^*\big(11^{13^{19}}\big).$$

Levels of the mosaic of n. Following Mullin's work, Gillman [1990, 1992] defined new functions on the mosaic of n. He used the concept of *levels* of the mosaic to describe the different tiers of exponentiation.

Suppose

$$n = 2^{3\cdot 5^7}\cdot 17^{23^{19}\cdot 29^{13^{11}\cdot 89}},$$

then the 2 and 17 are on the first level, the 3, 5, 23 and 29 are on the second level, the 7, 19, 13, and 89 are on the third level, and the 11 is on the fourth level of the mosaic.

Using this idea, Gillman generalized the Möbius function as follows:

$$\mu_i(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if the mosaic of } n \text{ has duplicate primes in the first } i \text{ levels} \\ & \quad \text{(including multiplicities at the } i\text{-th level),} \\ (-1)^k & \text{if the mosaic of } n \text{ consists of } k \text{ distinct primes in the first} \\ & \quad i \text{ levels.} \end{cases}$$

With this definition,

$$\mu_\infty = \mu^* \quad \text{and} \quad \mu_1 = \mu.$$

Gillman also extended Mullin's concept of generalized multiplicative to include the levels of the mosaic.

Definition 1.2. A function f is *i-multiplicative* if and only if $f(mn) = f(m)f(n)$ when m and n have no primes in common in the first i levels of their mosaic.

Gillman proved μ_i is i-multiplicative for all i. He then extended Mullin's work on $\psi(n)$ by generalizing it to depend on the levels of the mosaic.

Definition 1.3. For a fixed i and j such that $j > i$, the function $\psi_{j,i}(n)$ is computed as follows: Expand n through the first j levels of its mosaic; for each prime p on the i-th level of this expansion, multiply p by the product of the primes in the $(i+1)$-th through j-th levels above p, including multiplicities of the primes at the j-th level.

The following examples illustrate these computations:

$$\psi_{6,3}\big(2^{3^{5^{7^{11\cdot 13^2}}}} \cdot 3^{2\cdot 5^{3\cdot 7}}\big) = 2^{3^{5\cdot 7\cdot 11\cdot 13\cdot 2}} \cdot 3^{2\cdot 5^{3\cdot 7}},$$

$$\psi_{\infty,1}(n) = \text{product of all primes in the mosaic} = \psi(n).$$

Gillman also introduced the concept of i-relatively prime mosaics. That is, two integers, m and n, are *i-relatively prime* when they have no primes in common in the first i levels of their mosaics. Thus, the integers with mosaics 2^{3^5} and 7^{11^3} are 2-relatively prime, but *not* 3-relatively prime.

Motivation. In this paper, we will introduce new families of functions on the mosaic of n and determine which of these are i-multiplicative or i-additive. In Section 2, the functions will depend only on the primes present in the first i levels and their multiplicities. In Section 3, we evaluate previous attempts to generalize the concept of a divisor to the mosaic, and in Section 4 we introduce a new definition of a mosaic divisor that we believe will be more useful.

2. Mosaic functions

The functions Ω, ω, and λ are number theoretic functions that can be easily generalized to the mosaic of n. We discuss their generalizations because they are either i-multiplicative or i-additive. We also introduce a new function, ψ^*, which is interesting since it is either i-multiplicative or i-additive depending on the value of i.

The functions Ω_i and ω_i. $\Omega(n)$ is the total number of primes in the factorization of n, including repetitions. We generalize this idea with the function $\Omega_i(n)$, the total number of primes in the first i levels of the mosaic of n, including multiplicities on the i-th level. Thus, as an example,

$$\Omega_2\big(2^{3^7}\cdot 5^{11^6}\big) = \Omega_2(2^{3\cdot 3\cdot 3\cdot 3\cdot 3\cdot 3\cdot 3}\cdot 5^{11\cdot 11\cdot 11\cdot 11\cdot 11\cdot 11}) = 15,$$
$$\Omega_3\Big(7^{5^{7^5}\cdot 19}\cdot 19^{13^{7\cdot 11^3}}\Big) = \Omega_3\big(7^{5^{7\cdot 7\cdot 7\cdot 7\cdot 7}\cdot 19}\cdot 19^{13^{7\cdot 11\cdot 11\cdot 11}}\big) = 14.$$

Gillman [1990] proved that if f is i-multiplicative then f is j-multiplicative for all $j \geq i$. This is because if there are no primes in common in the first j levels of m and n, then there will clearly be no primes in common in the first i levels of m and n when $j \geq i$. This property is used to prove the following theorem.

Theorem 2.1. *For all i, Ω_i is j-additive for all j.*

Proof. Let m and n be 1-relatively prime. Then $\Omega_i(mn)$ is summing the number of prime divisors of the product mn and the number of primes in levels two though i of the mosaic of mn, including multiplicities at the i-th level. Since m and n are 1-relatively prime, the first term of this sum can be written as the number of prime divisors of m plus the number of prime divisors of n. Similarly, the second term can be written as the number of primes in levels two through i of the mosaic of m plus the number of primes in levels two through i of the mosaic of n, including multiplicities at the i-th level in each of these sums. Rearranging these sums results in $\Omega_i(m)+\Omega_i(n)$. Thus Ω_i is 1-additive and therefore j-additive for all j. □

The following two examples illustrate that it is necessary and sufficient that the first level of the mosaics have distinct primes in order that Ω_i be i-additive, as suggested by the previous proof:

$$\Omega_2\big(2^{3^5}\cdot 3^3\big) = 8 = 6+2 = \Omega_2\big(2^{3^5}\big)+\Omega_2(3^3),$$
$$\Omega_2\big(3^{3^5}\cdot 3^3\big) = \Omega_2(3^{2\cdot 3\cdot 41}) = 4 \neq 8 = 6+2 = \Omega_2\big(3^{3^5}\big)+\Omega_2(3^3).$$

Similarly, $\omega(n)$, the number of distinct primes in the prime factorization of n, can be generalized as $\omega_i(n)$, the number of *distinct* primes in the first i levels of the mosaic of n.

Since

$$\omega_4\big(7^{11\cdot 13^{53^2}}\cdot 29^{17^{89^2}}\big) = 9 = 5+4 = \omega_4\big(7^{11\cdot 13^{53^2}}\big)+\omega_4\big(29^{17^{89^2}}\big)$$

and

$$\omega_4\big(11^{5\cdot 11}\cdot 7^{19^3\cdot 53^{5^7}}\big) = 6 \neq 2+5 = \omega_4(11^{5\cdot 11})+\omega_4\big(7^{19^3\cdot 53^{5^7}}\big)$$

as counterexamples, ω_i is not 1-additive and therefore not j-additive for all j. Rather, as the following theorem demonstrates, it is j-additive for $j \geq i$.

Theorem 2.2. *For all i, ω_i is j-additive for $j \geq i$.*

Proof. Let m and n be i-relatively prime. Then $\omega_i(mn)$ is summing the number of prime divisors of the product mn and the distinct primes in levels two though i of the mosaic of mn (which must also be distinct from the prime divisors). Since m and n are relatively prime, the first term of this sum can be written as the prime divisors of m plus the prime divisors of n. Similarly, the second term can be written as the number of distinct primes in levels two through i of the mosaic of m plus the number of distinct primes in levels two through i of the mosaic of n. Thus $\omega_i(mn) = \omega_i(m) + \omega_i(n)$ and therefore ω_i is i-additive. This implies that ω_i is j-additive for $j \geq i$. □

The function $\boldsymbol{\lambda_i}$. The Liouville function, $\lambda(n) = (-1)^{\Omega(n)}$, also generalizes easily in the obvious way as $\lambda_i(n) = (-1)^{\Omega_i(n)}$. This leads to the following theorem, again recalling that i-multiplicative implies j-multiplicative for $j \geq i$.

Theorem 2.3. *For all i, λ_i is j-multiplicative for all $j \geq i$.*

Proof. Assume m and n are 1-relatively prime. It follows that

$$\begin{aligned}\lambda_i(mn) &= (-1)^{\Omega_i(mn)} = (-1)^{\Omega_i(m)+\Omega_i(n)} \\ &= (-1)^{\Omega_i(m)}(-1)^{\Omega_i(n)} = \lambda_i(m)\lambda_i(n).\end{aligned}$$

λ_i is 1-multiplicative and therefore j-multiplicative for all j. □

The function $\boldsymbol{\psi^*_{j,i}}$. Mullin defined the function ψ as the product of all primes in a mosaic. Gillman later extended this to the levels of the mosaic by introducing the function $\psi_{j,i}$. Mullin also defined the function ψ^* as the sum of the primes in a mosaic. To generalize this idea to the levels of the mosaic as well, we define the function $\psi^*_{j,i}$.

Definition 2.4. For fixed i and j, such that $j \geq i$, compute $\psi^*_{j,i}(n)$ as follows: Expand n through the first j levels of its mosaic; for each prime p on the i-th level of this expansion, add p to the sum of the primes in the $(i+1)$-th through j-th levels above p, including the multiplicities of the primes at the j-th level, then convert multiplication on the i-th level to addition.

Again, two examples help to illustrate this computation:

$$\begin{aligned}\psi^*_{4,2}\big(17^{11^{3^7}\cdot 19} \cdot 23^{2^{3^5}}\big) &= 17^{11+3+7+19} \cdot 23^{2+3+5}, \\ \psi^*_{4,1}\big(3^{5^{2^3}\cdot 7}\big) &= 3+5+2+3+7 = 20.\end{aligned}$$

Similar to the previous functions, $\psi^*_{j,1}$ is 1-additive, and therefore:

Theorem 2.5. *$\psi^*_{j,1}(n)$ is k-additive for all j and k.*

Proof. Let m and n be integers which are 1-relatively prime. $\psi^*_{j,1}(mn)$ is the sum of primes in the first j levels of the mosaic of mn, including multiplicities on the j-th level. This is equivalent to the sum of prime divisors of mn plus the sum of the primes in levels two through j of the mosaic of mn including multiplicities at the j-th level. Since m and n are relatively prime, the sum of prime divisors of mn can be written as the sum of prime divisors of m plus the sum of prime divisors of n. Similarly, the second term can be written as the sum of primes in levels two through j of the mosaic of m plus the number of primes in levels two through j of the mosaic of n including multiplicities at the j-th level in each. Rearranging these sums results in $\psi^*_{j,1}(m)+\psi^*_{j,1}(n)$. Thus $\psi^*_{j,1}$ is 1-additive and therefore k-additive for all k. □

Interestingly, while $\psi^*_{j,i}$ is k-additive for all j and k when $i=1$, for any $i>1$, $\psi^*_{j,i}$ is 1-multiplicative and therefore k-multiplicative for all j and k.

Theorem 2.6. *For all $i>1$ and $j\geq i$, $\psi^*_{j,i}(n)$ is k-multiplicative for all k.*

Proof. Let m and n be 1-relatively prime integers with prime factorizations

$$p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_r^{\alpha_r} \quad \text{and} \quad q_1^{\beta_1}q_2^{\beta_2}\cdots q_s^{\beta_s},$$

respectively. Because m and n are 1-relatively prime,

$$mn = p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_r^{\alpha_r}q_1^{\beta_1}q_2^{\beta_2}\cdots q_s^{\beta_s}.$$

Further, because $i>1$, $\psi^*_{j,i}(mn)$ has the same first $(i-1)$ levels as the mosaic of mn and the i-th level is equal to the i-th level of $\psi_{j,i}(mn)$ with multiplication converted to addition. Thus the unchanged first level can be partitioned into the parts that have the same first $(i-1)$ levels as m and n and with the i-th levels equal to the i-th levels of $\psi_{j,i}(m)$ and $\psi_{j,i}(n)$ respectively with multiplication converted to addition. That is,

$$\begin{aligned}\psi^*_{j,i}(mn) &= p_1^{a_1}p_2^{a_2}\cdots p_r^{a_r}q_1^{b_1}q_2^{b_2}\cdots q_s^{b_s}\\ &= (p_1^{a_1}p_2^{a_2}\cdots p_r^{a_r})(q_1^{b_1}q_2^{b_2}\cdots q_s^{b_s})\\ &= \psi^*_{j,i}(m)\psi^*_{j,i}(n),\end{aligned}$$

where a_* and b_* are the second through i-th levels of the mosaic, with the $(i+1)$-th to j-th levels brought down to the i-th level with multiplication converted to addition. Thus $\psi^*_{j,i}(mn)$ is 1-multiplicative and therefore k-multiplicative for all k. □

3. Early attempts at mosaic divisors

Many number theoretic functions are defined in terms of the divisors of n, so an analogous concept is needed for the mosaic. In this section, we examine two early attempts at this.

Submosaics. A mosaic can be viewed as a connected graph where the primes in the mosaic are the vertices and there is an edge between vertices if one prime is multiplied by the other or one is an exponent of the other. Mullin [1965] introduced the concept of submosaics as the mosaic corresponding to a connected subgraph of the graph of the full mosaic. Therefore, submosaics seems like a natural candidate for a mosaic divisor.

Mullin tried to show that functions of the form

$$F(n) = \sum f(d),$$

where the sum is over the set of submosaics of n, are generalized multiplicative when f is generalized multiplicative. Unfortunately, this is not true. If we let $C(n)$ be the set of all submosaics of n, Mullin assumed that $C(mn) = C(m) \times C(n)$, but this is not true as shown in the example:

$$2 \in C(13^2), \quad 17 \in C(17^5), \quad 2 \cdot 17 \notin C(13^2 \cdot 17^5).$$

Givisors. The givisor, from Gillman's divisor, was Gillman's attempt to generalize the concept of a divisor for mosaics. We examine this concept and its implications in this subsection and the next.

Definition 3.1. Let $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$. Each $p_i^{\alpha_i}$ is a *prime* givisor. Then a givisor of n is

(a) 1,

(b) a prime givisor, or

(c) a product of 1-relatively prime givisors from part (b).

We denote the set of all givisors of n by $G(n)$ and, as an example, consider

$$n = 2^{3^5} \cdot 3^{5^{17}} \cdot 5.$$

The givisors of n are

$$G\big(2^{3^5} \cdot 3^{5^{17}} \cdot 5\big) = \big\{1,\ 2^{3^5},\ 3^{5^{17}},\ 5,\ 2^{3^5} \cdot 3^{5^{17}},\ 2^{3^5} \cdot 5,\ 3^{5^{17}} \cdot 5,\ 2^{3^5} \cdot 3^{5^{17}} \cdot 5\big\}.$$

Gillman selected this structure because the mosaic above each prime in the first level is fixed, and the structure of the remaining mosaic does not change when a mosaic is divided by a givisor; we are simply splitting the mosaic into two parts.

In particular, givisors solve the fundamental problem that submosaics have, as we see in the following lemma.

Lemma 3.2. *For all i, $G(mn) = G(m) \times G(n)$ when m and n are i-relatively prime.*

Proof. Let the prime-power factorizations of m and n be

$$p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s} \quad \text{and} \quad q_1^{b_1} q_2^{b_2} \cdots q_t^{b_t},$$

respectively. Since m and n are i-relatively prime, the set of primes in the first level of m and the set of primes in the first level of n have no common elements. Therefore, the prime-power factorization of mn is

$$mn = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s} q_1^{b_1} q_2^{b_2} \cdots q_t^{b_t}.$$

If $d \in G(mn)$, then

$$d = p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s} q_1^{f_1} q_2^{f_2} \cdots q_t^{f_t}$$

where e_i is either 0 or a_i for $i = 1, 2, \ldots, s$ and f_j is either 0 or b_j for $j = 1, 2, \ldots, t$. Now let

$$d_1 = \gcd(d, m) \quad \text{and} \quad d_2 = \gcd(d, n).$$

Then

$$d_1 = p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s} \quad \text{and} \quad d_2 = q_1^{f_1} q_2^{f_2} \cdots q_t^{f_t}.$$

It follows that $d_1 \in G(m)$ and $d_2 \in G(n)$. Since $d = d_1 d_2$, $d \in G(m) \times G(n)$

Similarly, if $d_1 \in G(m)$ and $d_2 \in G(n)$, then $d_1 d_2 \in G(m) \times G(n)$ and $d_1 d_2 \in G(mn)$. Thus the sets are the same. □

Theorem 3.3. *If f is an i-multiplicative function, then*

$$F(n) = \sum_{d \in G(n)} f(d)$$

is also i-multiplicative.

Proof. To show that F is an i-multiplicative function, we must show that when m and n are i-relatively prime, $F(mn) = F(m)F(n)$. So assume that m and n are i-relatively prime. We know

$$F(mn) = \sum_{d \in G(mn)} f(d).$$

By the lemma, each givisor of mn can be written as the product $d = d_1 d_2$ where $d_1 \in G(m)$ and $d_2 \in G(n)$, and d_1 and d_2 are i-relatively prime. So

$$F(mn) = \sum_{d_1 \in G(m), d_2 \in G(n)} f(d_1 d_2).$$

Because f is i-multiplicative, and d_1 and d_2 are i-relatively prime, we see that

$$F(mn) = \sum_{d_1 \in G(m)} \sum_{d_2 \in G(n)} f(d_1) f(d_2) = \sum_{d_1 \in G(m)} f(d_1) \sum_{d_2 \in G(n)} f(d_2) = F(m)F(n). \quad \square$$

Functions defined by givisors. Givisors provide a mechanism for generalizing functions dependent on the concept of a divisor, and in this subsection we generalize three of these: τ — the number of divisors of n, σ — the sum of the positive divisors of n, and ϕ — the number of integers less than n relatively prime to n.

The function ${}_g\tau(n)$ counts the number of givisors of n, and hence can be computed by the formula

$${}_g\tau(n) = \sum_{d \in G(n)} 1.$$

The value of ${}_g\tau$ will always be a power of two with the exponent equal to the number of prime divisors of n. Using Theorem 3.3 with $f(d) = 1$, which is obviously i-multiplicative, we obtain:

Corollary 3.4. *${}_g\tau$ is i-multiplicative for all i.*

Similarly, we define ${}_g\sigma(n)$ as the sum of the givisors of n, and compute it using the formula

$${}_g\sigma(n) = \sum_{d \in G(n)} d.$$

Again using Theorem 3.3, except with $f(d) = d$, which is also i-multiplicative, we have:

Corollary 3.5. *${}_g\sigma$ is i-multiplicative for all i.*

We can generalize the concept of the number theoretic function $\phi(n)$, the number of integers less than n relatively prime to n, but the canonical formula for computing this,

$$\phi(n) = \sum_d \mu(d) \frac{n}{d},$$

does not generalize with it. Thus, while we can define $\phi_i(n)$ as the number of integers less than n that are i-relatively prime to n, it is not computed by the obvious generalization of the ϕ function, as given here:

$${}_gh_i(n) = \sum_{d \in G(n)} \mu_i(d) \frac{n}{d}.$$

By letting $n = 2^3$, we can compute ${}_gh_2(n) = 9$, but also determine, by listing the integers, that the number of integers 2-relatively prime to 8 is only 3. In spite of this significant disappointment, we have:

Theorem 3.6. *${}_gh_i$ is i-multiplicative for all i.*

Proof.
$$
\begin{aligned}
{}_g h_i(mn) &= \sum_{d\in G(mn)} \mu_i(d)\frac{mn}{d} = \sum_{d\in G(m)\times G(n)} \mu_i(d)\frac{mn}{d}\\
&= \sum_{d_1\in G(m)}\sum_{d_2\in G(n)} \mu_i(d_1)\mu_i(d_2)\frac{m}{d_1}\frac{n}{d_2}\\
&= \sum_{d_1\in G(m)} \mu_i(d_1)\frac{m}{d_1}\sum_{d_2\in G(n)} \mu_i(d_2)\frac{n}{d_2} = {}_g h_i(m)\,{}_g h_i(n). \qquad \square
\end{aligned}
$$

It is worth noticing that givisors are defined independently of the levels of the mosaic; that is, each positive integer n has the same set of givisors no matter which level i that we consider. Thus, the values of ${}_g\tau$ and ${}_g\sigma$ do not change as i varies, and ${}_g h_i$ only varies with i because μ_i changes as i varies.

To make these functions more dependent on the level i of the mosaic being considered, we might compose them with $\psi_{i,1}$, and examine functions of the form

$$ {}_g f_i(n) = {}_g f \circ \psi_{i,1}(n). $$

These functions are also i-multiplicative if ${}_g f$ is, and do vary in value with the choice of i.

If the integer n is squarefree, then

$$ {}_g h_i \circ \psi_{i,1}(n) = {}_g h_i(n) = \phi(n). $$

Further, the function

$$ {}_g\tau \circ \psi_{i,1}(n) $$

will always result in a power of two, but in this case the exponent is equal to the number of *distinct* primes in the first i levels of n. Hence, another formula is

$$ {}_g\tau \circ \psi_{i,1}(n) = 2^{\omega_i(n)}. $$

4. Mivisors

Neither submosaics nor givisors capture the properties of divisors that are desired. Submosaics do not effectively partition mosaics, and givisors are not sensitive to the parameter i representing the levels of the mosaic. With these two concerns in mind, we turn our investigation to a more promising generalization.

Definition of a mivisor. We define *prime* i-mivisors, mosaic divisor, as follows.

Definition 4.1. Let $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$. For each p_j, a *prime i-mivisor of n* is $p_j^{\alpha_j}$ expanded through i levels with multiplicities above the i-th level truncated, denoted $P_{j,i}$.

Let

$$ P_{j,i}^{(a_{j,1},a_{j,2},\cdots,a_{j,s_j})} $$

denote $p_j^{\alpha_j}$ expanded through i levels including the multiplicities $a_{j,1}, a_{j,2}, \ldots, a_{j,s_j}$ on the $(i+1)$-th level. For example, if

$$n = 2^{2^4 3^7} \cdot 5^{11},$$

then

$$P_{1,2} = 2^{2\cdot 3} \quad \text{and} \quad P_{1,2}^{(a_{1,1}, a_{1,2}, \cdots, a_{1,s_1})} = P_{1,2}^{(4,7)} = 2^{2^4 3^7}.$$

With this notation, we obtain the following definition.

Definition 4.2. If $n = \prod_{j=1}^{k} P_{j,i}^{(a_{j,1}, a_{j,2}, \cdots, a_{j,s_j})}$, then an i-mivisor of n is

(a) 1,
(b) $P_{j,i}^{(b_{j,1}, b_{j,2}, \cdots, b_{j,s_j})}$ where $1 \le b_j \le a_j$, or
(c) a product of 1-relatively prime i-mivisors from part (b).

We denote the set of all i-mivisors of n by $M_i(n)$ and, as an example, consider

$$n = 2^{3^{5^3\cdot 7}\cdot 5} \cdot 3^{7^{11^2}}.$$

The prime 3-mivisors of n are $2^{3^{5\cdot 7}\cdot 5}$ and $3^{7^{11}}$, and the set $M_3(n)$ is

$$\Big\{1,\ 2^{3^{5\cdot 7}\cdot 5},\ 2^{3^{5^2\cdot 7}\cdot 5},\ 2^{3^{5^3\cdot 7}\cdot 5},\ 3^{7^{11}},\ 3^{7^{11^2}},\ 2^{3^{5\cdot 7}\cdot 5}\cdot 3^{7^{11}},\ 2^{3^{5^2\cdot 7}\cdot 5}\cdot 3^{7^{11}},$$
$$2^{3^{5^3\cdot 7}\cdot 5}\cdot 3^{7^{11}},\ 2^{3^{5\cdot 7}\cdot 5}\cdot 3^{7^{11^2}},\ 2^{3^{5^2\cdot 7}\cdot 5}\cdot 3^{7^{11^2}},\ 2^{3^{5^3\cdot 7}\cdot 5}\cdot 3^{7^{11^2}}\Big\}.$$

We immediately have the following lemma.

Lemma 4.3. *For all i, $M_i(mn) = M_i(m) \times M_i(n)$.*

Proof. Let m and n be i-relatively prime integers such that $m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_t^{\alpha_t}$ and $n = q_1^{\beta_1} q_2^{\beta_2} \cdots q_u^{\beta_u}$. After applying the FTA to generate i levels of the mosaics of m and n, let

$$m = \prod_{j=1}^{t} P_{j,i}^{(a_{j,1}, a_{j,2}, \cdots, a_{j,r_j})} \quad \text{and} \quad n = \prod_{j=1}^{u} Q_{j,i}^{(b_{j,1}, b_{j,2}, \cdots, b_{j,s_j})}.$$

Since m and n are i-relatively prime,

$$mn = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_t^{\alpha_t} q_1^{\beta_1} q_2^{\beta_2} \cdots q_u^{\beta_u} = \prod_{j=1}^{t} P_{j,i}^{(a_{j,1}, a_{j,2}, \cdots, a_{j,r_j})} \prod_{j=1}^{u} Q_{j,i}^{(b_{j,1}, b_{j,2}, \cdots, b_{j,s_j})}.$$

If d is an i-mivisor of mn, then

$$d = \prod_{k=1}^{v} P_{k,i}^{\prime(c_{k,1}, c_{k,2}, \cdots, c_{k,r_k})} \prod_{k=1}^{w} Q_{k,i}^{\prime(e_{k,1}, e_{k,2}, \cdots, e_{k,s_k})},$$

where for each k, $P'_{k,i} = P_{j,i}$ for some j and $1 \le c_{k,\ell} \le a_{j,\ell}$ and $Q'_{k,i} = Q_{j,i}$ for some j and $1 \le e_{k,\ell} \le b_{j,\ell}$. Let d_1 be an i-mivisor of m such that

$$d_1 = \prod_{k=1}^{v} P_{k,i}^{\prime(c_{k,1},c_{k,2},\cdots,c_{k,r_k})}.$$

Let d_2 be an i-mivisor of n such that

$$d_2 = \prod_{k=1}^{w} Q_{k,i}^{\prime(e_{k,1},e_{k,2},\cdots,e_{k,s_k})}.$$

It follows that $d_1 \in M_i(m)$ and $d_2 \in M_i(n)$. Then d_1 and d_2 are i-relatively prime and $d = d_1 d_2$, so $d \in M_i(m) \times M_i(n)$.

Similarly, if $d_1 \in M_i(m)$ and $d_2 \in M_i(n)$, then $d_1 d_2 \in M_i(m) \times M_i(n)$ and $d_1 d_2 \in M_i(mn)$. Therefore the sets are the same. $\square$

Theorem 4.4. *If f is an i-multiplicative function, then*

$$F(n) = \sum_{d \in M_i(n)} f(d)$$

is also i-multiplicative.

Proof. Similar to Theorem 3.3. $\square$

The functions ${}_m\tau_i$ and ${}_m\sigma_i$. Similar to previous section, we let ${}_m\tau_i$ count the number of i-mivisors of n, and it is therefore computed as

$${}_m\tau_i(n) = \sum_{d \in M_i(n)} 1.$$

By Theorem 4.4, we find

Corollary 4.5. *For all i, ${}_m\tau_i$ is i-multiplicative.*

${}_m\tau_i(n)$ can be computed easily, as we see in the following lemma and theorem.

Lemma 4.6. *Let p be a prime and α be a positive integer. Then*

$${}_m\tau_i(p^\alpha) = 1 + \prod a_j,$$

where a_j is an element of the unfactored $(i+1)$ level of $p^\alpha = P_{\cdot,i}^{(a_1,a_2,\ldots,a_k)}$.

Proof. 1 is an i-mivisor of p^α and so is $P_{\cdot,i}^{(b_1,b_2,\ldots,b_k)}$ where $1 \le b_j \le a_j$ for all j. Since there are $\prod a_j$ ways to select the set $\{b_1, b_2, \ldots, b_j\}$, there are $1 + \prod a_j$ i-mivisors of p^α. $\square$

Theorem 4.7. *Let n have the prime factorization $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_s^{\alpha_s}$. Then*

$$ {}_m\tau_i(n) = \prod_{k=1}^{s} \Big(1 + \prod a_j\Big), $$

where a_j is an element of the unfactored $(i+1)$ level of $p_k^{\alpha_k}$.

Proof. Because ${}_m\tau_i$ is i-multiplicative for all i, we see that

$$ {}_m\tau_i(n) = {}_m\tau_i(p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_s^{\alpha_s}) = {}_m\tau_i(p_1^{\alpha_1})\,{}_m\tau_i(p_2^{\alpha_2}) \cdots {}_m\tau_i(p_s^{\alpha_s}). $$

Inserting the values from Lemma 4.6, we see that

$$ {}_m\tau_i(n) = \prod_{k=1}^{s} \Big(1 + \prod a_j\Big). \qquad \square $$

Moving forward, we let the function ${}_m\sigma_i(n)$ sums the i-mivisors of n,

$$ {}_m\sigma_i(n) = \sum_{d \in M_i(n)} d. $$

By using Theorem 4.4 again, we obtain:

Corollary 4.8. *${}_m\sigma_i$ is i-multiplicative.*

The function ${}_m\phi_i$. Using i-mivisors, we can generalize the concept of a common i-mivisor of two integers in the obvious way and more importantly, generalize the notion of a greatest common divisor.

Definition 4.9. A mosaic d is the *greatest common i-mivisor of m and n*, when at least one of them is not 0, if all of these conditions are satisfied:

(a) d is positive;

(b) d is an i-mivisor of a and b;

(c) if c is an i-mivisor of a and b, then c is an i-mivisor of d.

We write the greatest common i-mivisor of m and n as $\mathrm{GCM}_i(m,n)$, and have the following examples:

$$ \mathrm{GCM}_3\big(2^{3^5},\ 2 \cdot 5^{3^2}\big) = 1, \qquad \mathrm{GCM}_3\big(2^{3^{5^{20}}} \cdot 3^{7^{11}},\ 2^{3^{5^{10}}} \cdot 7\big) = 2^{3^{5^{10}}}. $$

Finally, we say that two integers m and n are GCM_i relatively prime if and only if $\mathrm{GCM}_i(m,n) = 1$.

We can now generalize the ϕ function in a natural way, by letting ${}_m\phi_i(n)$ be the number of integers GCM_i relatively prime to n that are less than or equal to n. Notice that this is very different from the function ϕ_i, which counts the number of integers less than n that are i-relatively prime to n. The latter function only detects and responds to the presence of primes in the mosaic, whereas the former function is sensitive to both the presence and configuration of the primes in the array. Thus, 2^3 is not 2-relatively prime to 3^2, but they are GCM_2 relatively prime.

Unfortunately, the obvious generalization of the summation formula for ϕ does not compute ${}_m\phi_i$ and, worse still, ${}_m\phi_i(n)$ is *not* an i-multiplicative function, as we see in this final example:

$$ {}_m\phi_2(2) = 1, \quad {}_m\phi_2(3) = 2, \quad {}_m\phi_2(2 \cdot 3) = 3 \neq {}_m\phi_2(2) \cdot {}_m\phi_2(3). $$

5. Conclusion

In conclusion, we have generalized several number theoretic functions in terms of the levels of the mosaic and explored their properties, building on the work of Mullin and Gillman. Further, we refined the notion of a divisor for mosaics so that we could begin to look at a broader class of number theoretic functions and to develop an arithmetic for mosaics. However, there are still significant open problems, and the first among these is to look for ways to compute ϕ_i and ${}_m\phi_i$.

References

[Gillman 1990] R. A. Gillman, "Some new functions on the mosaic of n", *J. Natur. Sci. Math.* **30**:1 (1990), 47–56. MR 93a:11005 Zbl 0702.11004

[Gillman 1992] R. A. Gillman, "k-distributive functions on the mosaic of n", *J. Nat. Sci. Math.* **32** (1992), 25–29. Zbl 0768.11003

[Mullin 1964] A. A. Mullin, "On a final multiplicative formulation of the fundamental theorem of arithmetic", *Z. Math. Logik Grundlagen Math.* **10** (1964), 159–161. MR 28 #3000 Zbl 0163.03904

[Mullin 1965] A. A. Mullin, "A contribution toward computable number theory", *Z. Math. Logik Grundlagen Math.* **11** (1965), 117–119. MR 33 #1276 Zbl 0254.10008

[Mullin 1967a] A. A. Mullin, "On Möbius' function and related matters", *Amer. Math. Monthly* **74** (1967), 1100–1102. MR 37 #143 Zbl 0155.08502

[Mullin 1967b] A. A. Mullin, "On new theorems for elementary number theory", *Notre Dame J. Formal Logic* **8** (1967), 353–356. MR 38 #4397 Zbl 0189.04206

Received: 2008-02-08 Accepted: 2008-07-20

kbildh01@saintmarys.edu Saint Mary's College, Notre Dame, IN 46556, United States

jaredErickson2012@u.northwestern.edu
School of Engineering and Applied Science,
Northwestern University, 2145 Sheridan Road Room C210,
Evanston, IL 60208, United States

cara.tacoma@trnty.edu Trinity Christian College, Palos Heights, IL 60463,
United States

rick.gillman@valpo.edu Valparaiso University, Valparaiso, IN 46383, United States

INVOLVE 2:1(2009)

The most general planar transformations that map parabolas into parabolas

Michael Bolt, Timothy Ferdinands and Landon Kavlie

(Communicated by Michael Dorff)

Consider the space of vertical parabolas in the plane interpreted generally to include nonvertical lines. It is proved that an injective map from a closed region bounded by one such parabola into the plane that maps vertical parabolas to other vertical parabolas must be the composition of a Laguerre transformation with a nonisotropic dilation. Here, a Laguerre transformation refers to a linear fractional or antilinear fractional transformation of the underlying dual plane.

1. Introduction

A familiar result in complex analysis is that in the extended complex plane, $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the Möbius transformations map circles and lines to other circles and lines. In a beautiful paper from 1937, reprinted in [Blair 2000], Carathéodory proved the following converse result.

Theorem 1 ([Carathéodory 1937]). *Every arbitrary one-to-one correspondence between the points of a circular disc C and a bounded point set C' by which circles lying completely in C are transformed into circles lying in C' must always be either a direct or inverse transformation of Möbius.*

So not only are the circle preserving maps of $\widehat{\mathbb{C}}$ the Möbius transformations, but even locally, these are the only transformations that can map circles to circles.

In this paper we consider the analogous problem for the extended dual plane, $\widehat{\mathbb{D}} = \mathbb{D} \cup L_\infty$, where

$$\mathbb{D} = \{z = x + jy : x, y \in \mathbb{R},\ j^2 = 0\} \quad \text{and} \quad L_\infty = \{(\alpha j)^{-1} : \alpha \in \mathbb{R}\}.$$

Here, the linear fractional transformations are the Laguerre transformations. They map vertical parabolas and nonvertical lines to other vertical parabolas and nonvertical lines. We prove the following.

MSC2000: 51B15.

Keywords: dual number, Laguerre transformation, parabola.

All three authors were supported by the National Science Foundation under Grant No. DMS-0702939.

Theorem 2. *Every injective map from a closed region bounded by a vertical parabola or nonvertical line that maps vertical parabolas and nonvertical lines to vertical parabolas and nonvertical lines is the composition of a nonisotropic dilation* $d_\lambda : (x, y) \to (\lambda x, \lambda^2 y)$, $0 \neq \lambda \in \mathbb{R}$, *with a direct or indirect Laguerre transformation.*

This theorem arose from work that two of the authors did to solve a Beckman–Quarles-type theorem for the dual plane. In particular, they proved:

Theorem 3 ([Ferdinands and Kavlie 2009]). *Suppose T is a bijective mapping on the space of vertical parabolas so that for vertical parabolas A, B,*

$$\delta(A, B) = 1 \text{ if and only if } \delta(T(A), T(B)) = 1.$$

Then T is induced by a Laguerre transformation of the dual plane.

Here, δ is the distance between (intersecting) parabolas and is measured as the difference in slopes at their point of intersection. An important step in the proof of this theorem uses a simpler version of Theorem 2 that has a very different proof.

We mention that planar Laguerre geometry often refers to a geometry of oriented circles where distance is measured as the length of the common tangent. The connection to dual numbers is made clear in Yaglom [1968]. We mention, too, that the transformations we call Laguerre transformations can also be interpreted as parabolic Möbius transformations. See, for instance, the recent survey by Kisil [2007].

2. Geometry in the extended dual plane

Here we summarize the properties and geometry of the dual numbers. A comprehensive account is given by Yaglom [1968].

A dual number $z \in \mathbb{D}$ is a formal expression $z = x + jy$ where $x, y \in \mathbb{R}$ and $j^2 = 0$. These numbers form a commutative algebra over $\mathbb{R}$ where addition and multiplication are done in the obvious way. One identifies dual numbers with points in the real plane via $x + jy \in \mathbb{D} \leftrightarrow (x, y) \in \mathbb{R}^2$, just as in the case of complex numbers. The coordinates of $z = x + jy$ are the real part and dual part, respectively. So $x = \text{Real}\,(x + jy)$ and $y = \text{Dual}\,(x + jy)$. Figure 1 illustrates the geometry of the dual plane. In particular, addition in $\mathbb{D}$ is done by adding position vectors. Multiplication is done by multiplying the real parts and adding the slopes of the position vectors. Because of this, the modulus and argument of z are usually defined by $|z| \overset{\text{def}}{=} |x|$ and $\arg z \overset{\text{def}}{=} y/x$.

The direct and indirect Laguerre transformations are the linear fractional and antilinear fractional transformations,

$$\mu(z) = \frac{az + b}{cz + d} \quad \text{and} \quad \mu(z) = \frac{a\bar{z} + b}{c\bar{z} + d},$$

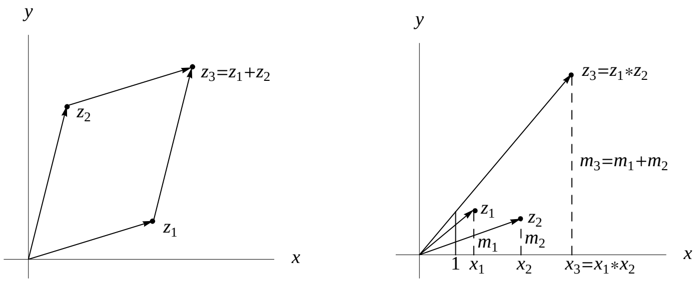


Figure 1. Addition and multiplication of dual numbers.

respectively, where $a, b, c, d \in \mathbb{D}$ and $ad - bc = 1$. The condition $ad - bc = 1$ acts as a normalization and has no effect on the transformation itself. It is just necessary that Real $(ad - bc) \neq 0$, or else μ maps $\widehat{\mathbb{D}}$ to a line or point.

The direct Laguerre transformations form a group that is isomorphic to $SL_2(\mathbb{D})$. Similar to Möbius transformations, they are generated by the following types:

(i) [translation] : $\mu(z) = z + b$ for $b \in \mathbb{D}$;

(ii) [rotation and isotropic dilation] : $\mu(z) = az$ for $a \in \mathbb{D}$, Real $(a) \neq 0$;

(iii) [inversion] : $\mu(z) = 1/z$.

These transformations preserve angles (measured as differences in slope); the indirect Laguerre transformations reverse angles.

Both the direct and indirect transformations preserve the space of vertical parabolas and nonvertical lines. (By a vertical parabola we mean that the axis of symmetry is vertical. The vertical parabolas and nonvertical lines can be described collectively as the graphs $y = rx^2 + sx + t$ for $r, s, t \in \mathbb{R}$.) This fact can be verified for the direct transformations by considering the three kinds of motions mentioned above. It then follows for indirect transformations, too, since obviously the conjugation $z \to \bar{z} = x - jy$ preserves the space.

By using stereographic projection, the extended dual plane $\widehat{\mathbb{D}} = \mathbb{D} \cup L_\infty$ can be viewed as an infinite cylinder as shown in Figure 2. (Laguerre transformations that do not arise as translations or similarities correspond with affine symmetries of the cylinder.) In this model, the set $L_\infty = \{(\alpha j)^{-1} : \alpha \in \mathbb{R}\}$ corresponds with a line of points at infinity. By using the transformation $\mu(z) = 1/z$ one can see that the parabola $y = rx^2 + sx + t$ intersects L_∞ at the point $-1/(rj)$, where $r, s, t \in \mathbb{R}$. In particular, nonvertical lines are the parabolas that intersect L_∞ at $1/(0j)$.

3. Proof of Theorem 2

The proof is modeled on Carathéodory's [Blair 2000]. We begin with an injective transformation $T : D \to D'$ that maps vertical parabolas and nonvertical lines

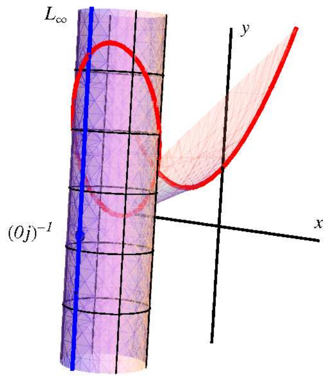


Figure 2. Representation of the dual plane on the Blaschke cylinder.

completely in D to other vertical parabolas and nonvertical lines in D'. Here, D and D' are regions in $\widehat{\mathbb{D}}$ bounded by a single vertical parabola or nonvertical line. By using a preliminary (direct or indirect) Laguerre transformation, we may assume that D is the closed region bounded above by the parabola $y = x^2$; that is, $D = \{y \le x^2\}$.

3.1. ***Preliminary remark.*** Using injectivity, it follows that the number of intersection points of two parabolas contained completely in D is preserved by T. For instance, if parabolas p_0 and p_1 intersect exactly once in D, and are therefore tangent, then the parabolas $T(p_0)$ and $T(p_1)$ also intersect exactly once in D', and are therefore tangent. (If they intersected twice, then one intersection point is the image of distinct points on p_0 and p_1.) Simply put, injectivity means that intersection points of parabolas cannot be created or destroyed.

3.2. ***First normalization.*** By postcomposing T with a direct or indirect Laguerre transformation, we may also assume:

(i) $T(0) = 0$,

(ii) $T\big((0j)^{-1}\big) = (0j)^{-1}$,

(iii) $T : \{y = 0\} \to \{y = 0\}$,

(iv) $T : \{y = x^2\} \to \{y = x^2\}$.

To see this, suppose that the original transformation is T_0, and $T_0(0) = w_0$ and $T_0\big((0j)^{-1}\big) = w_1$. If $\mu_1(w) = (w - w_0)/(w - w_1)$, then $T_1 \stackrel{\text{def}}{=} \mu_1 \circ T_0$ satisfies $T_1(0) = 0$ and $T_1\big((0j)^{-1}\big) = (0j)^{-1}$. We must determine a further Laguerre transformation μ_2 so that $T \stackrel{\text{def}}{=} \mu_2 \circ T_1$ satisfies (i)–(iv).

Since $T_1((0j)^{-1}) = (0j)^{-1}$, it already follows that T_1 maps nonvertical lines to nonvertical lines.

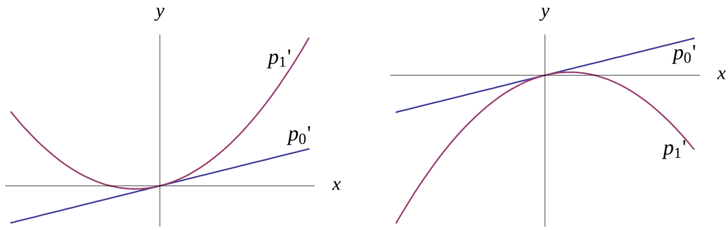


Figure 3. Intermediate configurations during normalization: cases (i) and (ii).

If $p_0 = \{y = 0\}$ and $p_1 = \{y = x^2\}$, then in fact, T_1 maps p_0 to a nonvertical line p_0' through the origin and p_1 to a parabola p_1' tangent to p_0' at the origin. See Figure 3 for the two possible cases. The set of Laguerre transformations that preserve 0 and $(0j)^{-1}$ have the form $\mu_2(z) = az$ or $\mu_2(z) = a\bar{z}$, where $a \in \mathbb{D}$ and Real $(a) \neq 0$.

For case (i), we use a direct transformation $\mu_2(z) = az$ where $\arg(a)$ is chosen so that $\mu_2(p_0') = \{y = 0\}$ and then $|a|$ is chosen so that $\mu_2(p_1') = \{y = x^2\}$. For case (ii), we use an indirect transformation $\mu_2(z) = a\bar{z}$ where the initial conjugation results in a configuration like case (i), and then a is chosen as just described. In both cases, $T \stackrel{\text{def}}{=} \mu_2 \circ T_1$ satisfies (i)–(iv), and we have exhausted our supply of Laguerre transformations.

3.3. ***Parallel lines.*** Recall that the nonvertical lines are exactly those parabolas that intersect L_∞ at the point $(0j)^{-1}$. After the first normalization, T preserves $(0j)^{-1}$, so it follows that T maps nonvertical lines to other nonvertical lines. Since T also preserves 0, it follows that T maps lines through the origin to other lines through the origin. (We used this fact in the second step of the normalization in Section 3.2.) Finally, parallel nonvertical lines intersect exactly once — the intersection occurs at $(0j)^{-1}$. Since T preserves this point, and since T preserves the number of intersection points among parabolas (Section 3.1), it follows that T maps parallel nonvertical lines to other parallel nonvertical lines. As a special case, T preserves $\{y = 0\}$, and so it also follows that T maps horizontal lines to horizontal lines.

3.4. ***Inscribed and circumscribed parabolas for a special unbounded polygon.*** It is a rather curious fact that the parabolas $y = x^2$ and $y = x^2 - 1/4$ arise as inscribed and circumscribed parabolas for a special unbounded polygon. The polygon is constructed from the lines tangent to the parabola $y = x^2$ at points that have integer valued coordinates. See Figure 4. Clearly, the parabola $y = x^2$ inscribes the polygon. To demonstrate that $y = x^2 - 1/4$ circumscribes the polygon, we show that

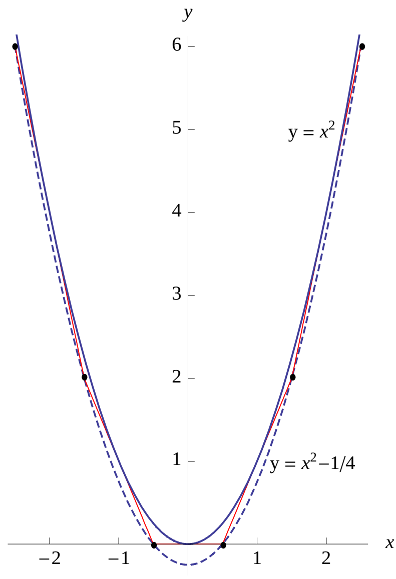


Figure 4. Inscribed and circumscribed parabolas for a special polygon.

the points of intersection of consecutive tangent lines to $y = x^2$ lie on its graph. The line tangent at (k, k^2) has equation $y = 2kx - k^2$. An easy calculation then shows that the lines tangent at (k, k^2) and $(k+1, (k+1)^2)$ intersect at $(k + 1/2, k^2 + k)$. Another easy calculation shows this point lies on $y = x^2 - 1/4$.

3.5. ***Second normalization.*** By further composing with a nonisotropic dilation $d_\lambda : (x, y) \to (\lambda x, \lambda^2 y)$, we assume $T(1, 1) = (1, 1)$. This is possible for the following reason. If after the first normalization the transformation is T_0, then $T_0(1, 1) = (\rho, \rho^2)$ for $0 \neq \rho \in \mathbb{R}$ since T_0 preserves $\{y = x^2\}$. If $\lambda = \rho^{-1}$ then $T \overset{\text{def}}{=} d_\lambda \circ T_0$ preserves $(1, 1)$. Furthermore, d_λ preserves 0, $(0j)^{-1}$, $\{y = 0\}$, and $\{y = x^2\}$. So T continues to satisfy conditions (i)–(iv) from Section 3.2.

Given that T maps horizontal lines to horizontal lines (Section 3.3) and now $T(1, 1) = (1, 1)$, it follows that T preserves $\{y = 1\}$. But T also preserves $\{y = x^2\}$, so it follows that T preserves both intersection points. In particular, $T(-1, 1) = (-1, 1)$.

Following the two normalizations (which identify the Laguerre transformation and nonisotropic dilation) we mention that the proof of Theorem 2 will be complete once we show that T is the identity transformation. We first prove that T reproduces the configuration in Figure 4.

To do this, notice that $y = x^2$ and $y = x^2 - 1/4$ are tangent parabolas — they intersect only at the point $-(1j)^{-1} \in L_\infty$. Since T preserves $\{y = x^2\}$, and since intersection points cannot be created or destroyed (Section 3.1), it follows that T transforms the parabola $y = x^2 - 1/4$ to a parabola $y = rx^2 + sx + t$ for $r, s, t \in \mathbb{R}$ with $s^2 - 4(r-1)t = 0$. (This is the required condition for a single intersection with $y = x^2$.)

Next, T maps tangent parabolas to tangent parabolas (Section 3.1) and nonvertical line segments to nonvertical line segments (Section 3.3). It also preserves $\{y = x^2\}$ as well as the points $(-1, 1)$, $(0, 0)$, and $(1, 1)$. It follows that the parabola $y = rx^2 + sx + t$ must contain the points of intersection of consecutive lines tangent to $y = x^2$ at (x_k, x_k^2) for a sequence of real numbers

$$\dots, x_{-2}, x_{-1} = -1, x_0 = 0, x_1 = 1, x_2, \dots.$$

We will verify that together with the required condition for single intersection, this demands $r = 1$, $s = 0$, $t = -1/4$, and $x_k = k$ for $k \in \mathbb{Z}$.

Notice that the line tangent to $y = x^2$ at (x_k, x_k^2) has equation $y = 2x_k x - x_k^2$. From this, one can check that the point of intersection of the lines tangent to $y = x^2$ at (x_k, x_k^2) and (x_{k+1}, x_{k+1}^2) is $\big((x_k + x_{k+1})/2, x_k x_{k+1}\big)$. Setting $k = -1$ and $k = 0$, this means the parabola $y = rx^2 + sx + t$ must contain points $(-1/2, 0)$ and $(1/2, 0)$. Together with the condition $s^2 - 4(r-1)t = 0$, this requires $r = 1$, $s = 0$, and $t = -1/4$. (The other possibility yields the parabola $y = 0$. This is ruled out by injectivity.)

At this point it is established that the parabola $y = x^2 - 1/4$ contains the intersection points $\big((x_k + x_{k+1})/2, x_k x_{k+1}\big)$. After some algebra, this can be restated as $1 = (x_{k+1} - x_k)^2$. So the x_k are evenly spaced with $x_k = k$ for $k = -1, 0, 1$. By injectivity, it follows that $x_k = k$ for $k \in \mathbb{Z}$.

3.6. ***Preservation of a dense subset of $y = x^2$.*** It follows from Section 3.5 that the normalized transformation T preserves the points

$$(k, k^2) \quad \text{for } k = 0, \pm 1, \pm 2, \pm 3, \dots.$$

Here we show as well that T preserves all points on $y = x^2$ whose coordinates are dyadic rational, that is, the coordinates have the form $k \cdot 2^{-q}$ for $k, q \in \mathbb{Z}$.

To do this, we construct another polygon like the one in Section 3.4. In particular, for fixed $q \geq 1$, we draw lines tangent to $y = x^2$ at the points $(k\,2^{-q}, k^2\,2^{-2q})$ for $k = 0, \pm 1, \pm 2, \pm 3, \dots$. It is a simple matter to check that the points of intersection of the consecutive tangent lines are $\big((2k+1)2^{-q-1}, k(k+1)2^{-2q}\big)$, and the parabola $y = x^2 - 2^{-2q-2}$ contains these intersection points. As in Section 3.5, T must transform this configuration to one consisting of the parabola $y = x^2$, lines tangent to $y = x^2$ at points (x_k, x_k^2) for a sequence of real numbers

$$\dots, x_{-2}, x_{-1}, x_0 = 0, x_1, x_2, \dots,$$

and a parabola $y = rx^2 + sx + t$ that has a single intersection with $y = x^2$. The parabola $y = rx^2 + sx + t$ must also contain the intersection points of the consecutive lines tangent to $y = x^2$ at the (x_k, x_k^2). Since T was normalized so that $T(1, 1) = (1, 1)$ and $T(-1, 1) = (-1, 1)$, we know that $x_{-2^q} = -1$ and $x_{2^q} = 1$. The claim will be proved if we show that $r = 1$, $s = 0$, $t = -2^{-2q-2}$, and $x_k = k\,2^{-q}$ for $k \in \mathbb{Z}$.

As things are arranged, $T(k\,2^{-q}, k^2\,2^{-2q}) = (x_k, x_k^2)$ for $k \in \mathbb{Z}$. Since T maps horizontal lines to horizontal lines, it follows that $x_k^2 = x_{-k}^2$, and in particular, $x_{-k} = -x_k$. This means that the target configuration must be symmetric with respect to the y-axis. Therefore, $s = 0$. Already, the single intersection of $y = x^2$ with $y = rx^2 + sx + t$ requires $s^2 - 4(r-1)t = 0$, so now $r = 1$ or $t = 0$. The case $t = 0$ is ruled out else T transforms the parabola $y = x^2 - 2^{-2q-2}$ to a parabola $y = rx^2$ that intersects $y = 0$ exactly once (if $r \neq 0$) or else infinitely many times (if $r = 0$).

We conclude that T transforms $y = x^2 - 2^{-2q-2}$ to a parabola $y = x^2 + t$ for some $0 \neq t \in \mathbb{R}$. In fact, since T preserves $\{y = 0\}$, it must be that $t < 0$. (This also uses Section 3.1.)

As in Section 3.5, the intersection points of the lines tangent to $y = x^2$ at (x_k, x_k^2) have the form $\big((x_k + x_{k+1})/2, x_k x_{k+1}\big)$ and they lie on $y = x^2 + t$. It follows that $x_k x_{k+1} = (x_k + x_{k+1})^2/4 + t$, or equivalently, $t = -(x_{k+1} - x_k)^2/4$ for $k \in \mathbb{Z}$. In particular, the x_k are evenly spaced. Since $x_0 = 0$ and $x_{2^q} = 1$, it follows that the distance from x_k to x_{k+1} is 2^{-q}, and therefore, $x_k = k\,2^{-q}$ for $k \geq 1$. (This also uses injectivity.) The same kind of argument applies for $k \leq -1$. Finally, one finds easily that $t = -(x_{k+1} - x_k)^2/4 = -2^{-2q-2}$.

3.7. ***Preservation of a dense subset of $y < x^2$.*** Guided by Carathéodory's argument [Carathéodory 1937, page 576], we now show that T preserves a set of points that is dense in $y < x^2$. To do this, we take all lines that are tangent to the parabola $y = x^2$ at points whose coordinates are dyadic rational. By the previous subsection, T preserves these points of tangency (Section 3.6) along with the parabola $y = x^2$ (Section 3.2), so it also preserves the lines tangent to $y = x^2$ at these points (Section 3.1, Section 3.3). It then follows that T preserves each point of intersection of these tangent lines. These intersection points form a dense subset of $y < x^2$.

The set includes, in particular, the points $(-.5, -2)$, $(0, -1)$, and $(+.5, -2)$. (They arise as the intersection points of the lines tangent at $x = -2, -1, +1, +2$.) Since T preserves these points, and since T maps vertical parabolas and nonvertical lines to vertical parabolas and nonvertical lines, it follows that T preserves the unique vertical parabola containing these points. In particular, T preserves the parabola $y = -4x^2 - 1$.

3.8. ***Completion of the proof of Theorem 2.*** We next show that T preserves each point of the parabola $y = x^2$.

Consider the alternative. Since T preserves $\{y = x^2\}$, the alternative is that there is a (nondyadic) real number b and $\tau \notin \{0, 1\}$ so that $T(b, b^2) = (\tau b, \tau^2 b^2)$. By once more replaying the arguments from Section 3.5–3.6, it would follow that $T(d \cdot b, d^2 \cdot b^2) = (d \cdot \tau b, d^2 \cdot \tau^2 b^2)$ for dyadic rational d. Moreover, T would map the line tangent to $y = x^2$ at $x = d \cdot b$ to the line tangent to $y = x^2$ at $x = d \cdot \tau b$.

Following the argument of Section 3.7, this determines how T would act on the set of points that arise as the intersection points of lines tangent to $y = x^2$ at $x = d \cdot b$ for d dyadic rational. On this dense set of points, $T(x, y) = (\tau x, \tau^2 y)$.

Consider now that d is a fixed (but arbitrary) dyadic rational number. The lines tangent to $y = x^2$ at $x = d \cdot b$ and $x = -2d \cdot b$ intersect at $(-d \cdot b/2, -2d^2 \cdot b^2)$, and the lines tangent to $y = x^2$ at $x = -d \cdot b$ and $x = 2d \cdot b$ intersect at $(+d \cdot b/2, -2d^2 \cdot b^2)$. These points determine the horizontal line $y = -2d^2 \cdot b^2$. As T maps horizontal lines to horizontal lines (Section 3.3), and since the intersection points belong to the set on which $T(x, y) = (\tau x, \tau^2 y)$, it follows that T would map the horizontal line $y = -2d^2 \cdot b^2$ to the horizontal line $y = -2d^2 \cdot \tau^2 b^2$.

Here lies the contradiction. If $\tau^2 < 1$, choose a dyadic d so that

$$(2b^2)^{-1} < d^2 < (2b^2 \tau^2)^{-1}.$$

Then the parabola $y = -4x^2 - 1$ intersects the line $y = -2d^2 \cdot b^2$ twice, but after the action of T, the parabola $y = -4x^2 - 1$ does not intersect the line $y = -2d^2 \cdot \tau^2 b^2$. This violates Section 3.1.

Similarly, if $\tau^2 > 1$, choose a dyadic d so that $(2b^2 \tau^2)^{-1} < d^2 < (2b^2)^{-1}$. Then the parabola $y = -4x^2 - 1$ does not intersect the line $y = -2d^2 \cdot b^2$, but after the action of T, the parabola $y = -4x^2 - 1$ intersects the line $y = -2d^2 \cdot \tau^2 b^2$. Again this violates Section 3.1.

There is the remaining case $\tau = -1$. For this we identify an asymmetric parabola preserved by T. For instance, the lines tangent to $y = x^2$ at $x = 0, 0.5, 2, 2.5, 4$ determine intersection points $(1, 0)$, $(2, 0)$, and $(1.5, 1.25)$. These points determine the parabola $y = -5x^2 + 15x - 10$ that is preserved by T. Next, we choose a dyadic d so that $|4bd - 5| < 5/\sqrt{3}$. This condition guarantees that the line tangent to $y = x^2$ at $x = d \cdot b$ does not intersect $y = -5x^2 + 15x - 10$, but the line tangent at $x = -d \cdot b$ does intersect $y = -5x^2 + 15x - 10$. Under the action of T, the tangent line at $x = d \cdot b$ would be mapped to the tangent line at $x = -d \cdot b$, creating an intersection with $y = -5x^2 + 15x - 10$. Again the contradiction.

We conclude that T preserves each point of $y = x^2$, and it follows from the argument in Section 3.7 that T preserves all points below $y = x^2$, since each such point can be expressed as the intersection of lines tangent to $y = x^2$. That is, T acts identically on $y \le x^2$, and following the remark in Section 3.5, this completes the proof of Theorem 2.

References

[Blair 2000] D. E. Blair, *Inversion theory and conformal mapping*, Student Mathematical Library **9**, American Mathematical Society, Providence, RI, 2000. MR 1779832 Zbl 0956.30001

[Carathéodory 1937] C. Carathéodory, "The most general transformations of plane regions which transform circles into circles", *Bull. Amer. Math. Soc.* **43** (1937), 573–579. Zbl 63.0294.03

[Ferdinands and Kavlie 2009] T. Ferdinands and L. Kavlie, "A Beckman–Quarles type theorem for Laguerre transformations in the dual plane", preprint, 2009.

[Kisil 2007] V. V. Kisil, "Starting with the group $SL_2(\mathbf{R})$", *Notices Amer. Math. Soc.* **54**:11 (2007), 1458–1465. MR 2361159

[Yaglom 1968] I. M. Yaglom, *Complex numbers in geometry*, Translated from the Russian by Eric J. F. Primrose, Academic Press, New York, 1968. MR 0220134

Received: 2008-09-05 Revised: Accepted: 2009-02-11

mbolt@calvin.edu
Department of Mathematics and Statistics, Calvin College, 1740 Knollcrest Circle SE, Grand Rapids, Michigan 49546-4403, United States
http://www.calvin.edu/~mdb7/

tdf3@calvin.edu
Department of Mathematics and Statistics, Calvin College, 1740 Knollcrest Circle SE, Grand Rapids, Michigan 49546-4403, United States

ljk7@calvin.edu
Department of Mathematics and Statistics, 1740 Knollcrest Circle SE, Calvin College, Grand Rapids, Michigan 49546-4403, United States

INVOLVE 2:1(2009)

Equidissections of kite-shaped quadrilaterals

Charles H. Jepsen, Trevor Sedberry and Rolf Hoyer

(Communicated by Kenneth S. Berenhaut)

Let $Q(a)$ be the convex kite-shaped quadrilateral with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, and (a, a), where $a > 1/2$. We wish to dissect $Q(a)$ into triangles of equal areas. What numbers of triangles are possible? Since $Q(a)$ is symmetric about the line $y = x$, $Q(a)$ admits such a dissection into any even number of triangles. In this article, we prove four results describing $Q(a)$ that can be dissected into certain odd numbers of triangles.

1. Introduction

We wish to dissect a convex polygon K into triangles of equal areas. A dissection of K into m triangles of equal areas is called an *m-equidissection*. The *spectrum* of K, denoted $S(K)$, is the set of integers m for which K has an m-equidissection. Note that if m is in $S(K)$, then so is km for all $k > 0$. If $S(K)$ consists of precisely the positive multiples of m, we write $S(K) = \langle m\rangle$ and call $S(K)$ *principal*.

Quite a bit is known about the spectrum of the trapezoid $T(a)$ with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(a, 1)$ for $a > 0$. For example, if a is rational with $a = r/s$, where r and s are relatively prime positive integers, then $S(T(a)) = \langle r + s\rangle$; if a is transcendental, then $S(T(a))$ is the empty set. See [Kasimatis and Stein 1990] or [Stein and Szabó 1994]. In addition, $S(T(a))$ is known for many irrational algebraic numbers a, particularly a satisfying a quadratic polynomial. See [Jepsen 1996; Jepsen and Monsky 2008; Monsky 1996]. For instance, if $a = (2r-1)+r\sqrt{3}$ where r is an integer ≥ 8, then $S(T(a)) = \{4r, 5r, 6r, \dots\}$.

Less is known about the spectrum of the kite-shaped quadrilateral $Q(a)$ with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, (a, a) for $a > 1/2$. Here certainly $S(Q(a))$ contains 2 and hence all even positive integers. If $a = 1$, then $Q(a)$ is a square, and in this case $S(Q(a)) = \langle 2\rangle$. See [Monsky 1970].) For other values of a, the question is, What odd numbers, if any, are in $S(Q(a))$? In Section 2, we prove four theorems that answer this question for certain a. In Section 3, we pose some questions that remain open.

MSC2000: 52B45.
Keywords: equidissection, spectrum.

2. Main results

As in the introduction, $Q(a)$ denotes the quadrilateral with vertices $(0,0)$, $(1,0)$, $(0,1)$, and (a,a) for $a > 1/2$. The following two results about $Q(a)$ are shown in [Kasimatis and Stein 1990, pages 290 and 291]:

(i) Let ϕ_2 be an extension to $\mathbb{R}$ of the 2-adic valuation on $\mathbb{Q}$. (See [Stein and Szabó 1994] for a discussion of valuations.) If $\phi_2(a) > -1$, then $S(Q(a)) = \langle 2 \rangle$. In particular, if a is transcendental, then $S(Q(a)) = \langle 2 \rangle$.

(ii) Let $a > 1/2$ be a rational number such that $\phi_2(a) \leq -1$. That is, $a = r/(2s)$, where r and s are relatively prime positive integers, r is odd, and $r > s$. Then $S(Q(a))$ contains all odd integers of the form $r + 2sk$ for $k \geq 0$.

[Kasimatis and Stein 1990] and [Stein and Szabó 1994] raise two questions:

- Are there rational numbers a with $\phi_2(a) \leq -1$ for which $S(Q(a))$ contains odd numbers less than r?
- Are there irrational algebraic numbers a with $\phi_2(a) \leq -1$ for which $S(Q(a))$ contains odd numbers? In particular, does $S(Q(\sqrt{3}/2))$ contain odd numbers?

We answer these questions in the affirmative. First we present a slight strengthening of statement (ii) above.

Theorem 1. *Let $a = r/(2s)$, where r and s are relatively prime positive integers, r is odd, and $r > s$. Then $S(Q(a))$ contains all integers of the form $r + 2k$ for $k \geq 0$.*

Proof. Partition $Q(a)$ into three triangles as in Figure 1, left. We want to find nonnegative integers t_1, t_2, t_3 so that the areas A_1, A_2, A_3 of the three triangles satisfy

$$A_1 t = a t_1, \qquad A_2 t = a t_2, \qquad A_3 t = a t_3, \tag{1}$$

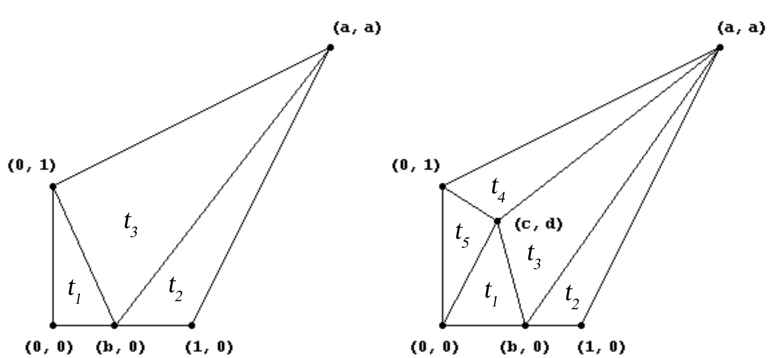


Figure 1

where $t = t_1 + t_2 + t_3$. (Note that the area of $Q(a)$ is a.) Then $Q(a)$ can be further dissected into t triangles each of area a/t. Here $A_1 = \frac{1}{2}b,\ \ A_2 = \frac{1}{2}a(1-b)$, and $A_3 = \frac{1}{2}(a + ab - b)$. For $k \geq 0$, choose $t_1 = s,\ \ t_2 = k,\ \ t_3 = r - s + k$, and $b = r/(r + 2k)$. Then $t = r + 2k$, $b = r/t$, and equations (1) are satisfied. Thus $r + 2k \in S(Q(a))$. □

Theorem 2. *Let a be as in Theorem 1, and suppose r is not a prime number. Then $S(Q(a))$ contains odd numbers less than r.*

Proof. We know that $S(Q(a)) = S(Q(a/(2a-1)))$ for any a [Kasimatis and Stein 1990, pages 284 and 285]. If $a = r/(2s)$, then $a/(2a-1) = r/((2(r-s))$. So replacing s by $r - s$ if necessary, we may assume s is odd. Partition $Q(a)$ into five triangles as shown in Figure 1, right. We want the areas A_1, A_2, A_3, A_4, A_5 of the triangles to satisfy

$$A_1 t = a t_1, \quad A_2 t = a t_2, \quad A_3 t = a t_3, \quad A_4 t = a t_4, \quad A_5 t = a t_5, \tag{2}$$

where $t = t_1 + t_2 + t_3 + t_4 + t_5$. In this case, $A_1 = \frac{1}{2}bd,\ \ A_2 = \frac{1}{2}a(1-b),\ \ A_5 = \frac{1}{2}c$, $A_4 = \frac{1}{2}(c(a-1) - a(d-1))$, and $A_3 = \frac{1}{2}(d(a-b) - a(c-b))$. Since r is an odd composite number, we can write $r = r_1 r_2$, where $3 \leq r_1 \leq r_2$.

Case (i): $s > r_2$. Choose $t_1 = 1,\ \ t_2 = \frac{1}{2}(s - r_1),\ \ t_3 = \frac{1}{2}(r_1 + r_2) - 1,\ \ t_4 = \frac{1}{2}(s - r_2)$, $t_5 = 0,\ \ b = r_1/s,\ \ c = 0$, and $d = r^2/s$. Then $t = s$, and we check that equations (2) are satisfied. Then $s \in S(Q(a))$ and $s < r$.

Case (ii): $s < r_2$. Choose $t_1 = \frac{1}{2}(r_1 - 1),\ \ t_2 = \frac{1}{2}(r_1 r_2 - r_1 - 2s),\ \ t_3 = \frac{1}{2}(r_2 + 1),\ \ t_4 = 0$, and $t_5 = \frac{1}{2}(r - r_2 - 2s)$. The assumption on s implies that the t_i are nonnegative, and their sum t is $r - 2s$. Now let $b = (t - 2t_2)/t = r_1/t,\ \ c = (2at_5)/t$, and $d = (2at_1)/(bt) = (2at_1)/r_1$. Then $s = tt_1 - r_1 t_5$, and again we check that equations (2) are satisfied. Thus $r - 2s \in S(Q(a))$ and $r - 2s < r$. □

Theorem 3. *Let $a = \sqrt{3}/2$. Then 21 is in $S(Q(a))$.*

Proof. Partition $Q(a)$ into five triangles shown in Figure 2, left. The areas of the five triangles are in the proportion

$$\frac{3}{14\sqrt{3}} : \frac{3}{14\sqrt{3}} : \frac{1}{14\sqrt{3}} : \frac{7}{14\sqrt{3}} : \frac{7}{14\sqrt{3}}$$

or $3:3:1:7:7$. Hence we can further dissect $Q(a)$ into $t = 3 + 3 + 1 + 7 + 7 = 21$ triangles each of area $1/(14\sqrt{3}) = (1/21)(\sqrt{3}/2)$. □

There are infinitely many radicals besides $\sqrt{3}/2$ that have odd numbers in their spectra. For example, the next theorem says $11 \in S(Q(\sqrt{5}/4))$, $15 \in S(Q(\sqrt{21}/4)$, $17 \in S(Q(\sqrt{33}/4)$, $21 \in S(Q(\sqrt{65}/4)$, and so forth.

Theorem 4. *For $k \geq 1$, let $a = \sqrt{(2k+1)(2k+3)}/(4\sqrt{3})$. Then $2k + 9$ lies in $S(Q(a))$.*

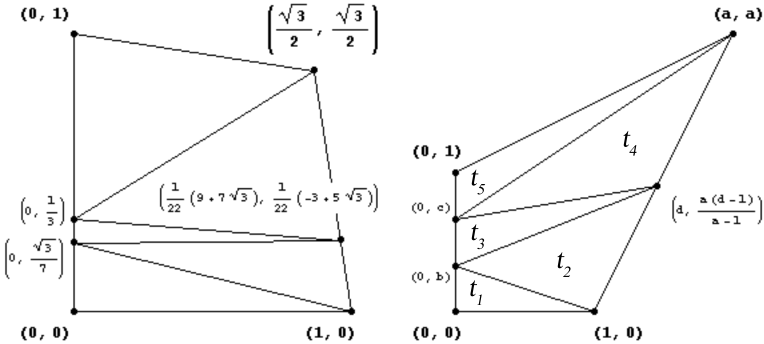


Figure 2

Proof. Partition $Q(a)$ into five triangles as shown in Figure 2, right. As before, we want the areas A_i of the triangles to satisfy equations (2) above. Here $A_1 = \frac{1}{2}b$, $A_3 = \frac{1}{2}(c-b)d$, $A_5 = \frac{1}{2}a(1-c)$,

$$A_2 = \frac{1}{2}\left(\frac{d-1}{a-1}\right)(a+ab-b), \quad \text{and} \quad A_4 = \frac{1}{2}\left(\frac{a-d}{a-1}\right)(a+ac-c).$$

Choose $t_1 = t_2 = t_3 = 2$, $t_5 = 3$, and $t_4 = 2k$, so $t = 2k+9$ and $48a^2 = (t-8)(t-6)$. Now let $b = (4a)/t$, $c = (t-6)/t$, and $d = (4a)/(t-6-4a)$. We show once again that equations (2) are satisfied. Thus $2k+9 \in S(Q(a))$. □

3. Open questions

While we have answered a few questions about odd numbers in $S(Q(a))$, many others remain:

(i) Is the converse of Theorem 2 true? That is, if a is as in Theorem 1 and r is a prime number, is r the smallest odd number in $S(Q(a))$?

(ii) Let a be as in Theorem 2. What is the smallest odd number in $S(Q(a))$? What are all the odd numbers in $S(Q(a))$?

(iii) Let a be an irrational algebraic number with $\phi_2(a) \leq -1$. Does $S(Q(a))$ always contain odd numbers?

(iv) Let a be arbitrary, and let m be an odd number. If m is in $S(Q(a))$, is $m+2$ in $S(Q(a))$? (This is the same as, Is $S(Q(a))$ closed under addition?)

Acknowledgment

This article constitutes part of the output of an undergraduate research project by Trevor Sedberry and Rolf Hoyer under the direction of Charles Jepsen.

References

[Jepsen 1996] C. H. Jepsen, "Equidissections of trapezoids", *Amer. Math. Monthly* **103**:6 (1996), 498–500. MR 97i:51031 Zbl 0856.51007

[Jepsen and Monsky 2008] C. H. Jepsen and P. Monsky, "Constructing equidissections for certain classes of trapezoids", *Discrete Math.* (2008).

[Kasimatis and Stein 1990] E. A. Kasimatis and S. K. Stein, "Equidissections of polygons", *Discrete Math.* **85**:3 (1990), 281–294. MR 91j:52017 Zbl 0736.05028

[Monsky 1970] P. Monsky, "On dividing a square into triangles", *Amer. Math. Monthly* **77** (1970), 161–164. MR 40 #5454 Zbl 0187.19701

[Monsky 1996] P. Monsky, "Calculating a trapezoidal spectrum", *Amer. Math. Monthly* **103** (1996), 500–501. MR 97i:51032 Zbl 0856.51008

[Stein and Szabó 1994] S. K. Stein and S. Szabó, *Algebra and tiling: Homomorphisms in the service of geometry*, Carus Mathematical Monographs **25**, Mathematical Association of America, Washington, DC, 1994. MR 95k:52035 Zbl 0930.52003

Received: 2007-06-10 Accepted: 2008-06-02

jepsen@math.grinnell.edu *Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112, United States*

sedberry@grinnell.edu *Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112, United States*

hoyerrol@grinnell.edu *Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112, United States*

Hamiltonian labelings of graphs

Willem Renzema and Ping Zhang

(Communicated by Ron Gould)

For a connected graph G of order n, the detour distance $D(u, v)$ between two vertices u and v in G is the length of a longest $u-v$ path in G. A Hamiltonian labeling of G is a function $c : V(G) \to \mathbb{N}$ such that $|c(u) - c(v)| + D(u, v) \geq n$ for every two distinct vertices u and v of G. The value $\mathrm{hn}(c)$ of a Hamiltonian labeling c of G is the maximum label (functional value) assigned to a vertex of G by c; while the Hamiltonian labeling number $\mathrm{hn}(G)$ of G is the minimum value of Hamiltonian labelings of G. Hamiltonian labeling numbers of some well-known classes of graphs are determined. Sharp upper and lower bounds are established for the Hamiltonian labeling number of a connected graph. The corona $\mathrm{cor}(F)$ of a graph F is the graph obtained from F by adding exactly one pendant edge at each vertex of F. For each integer $k \geq 3$, let $\mathscr{H}_k$ be the set of connected graphs G for which there exists a Hamiltonian graph H of order k such that $H \subset G \subseteq \mathrm{cor}(H)$. It is shown that $2k - 1 \leq \mathrm{hn}(G) \leq k(2k - 1)$ for each $G \in \mathscr{H}_k$ and that both bounds are sharp.

1. Introduction

The *distance* $d(u, v)$ between two vertices u and v in a connected graph G is the length of a shortest path between these two vertices. The *eccentricity* $e(v)$ of a vertex v in G is the maximum distance from v to a vertex of G. The *radius* $\mathrm{rad}(G)$ of G is the minimum eccentricity among the vertices of G, while the *diameter* $\mathrm{diam}(G)$ of G is the maximum eccentricity among the vertices of G. A vertex v with $e(v) = \mathrm{rad}(G)$ is called a *central vertex* of G. If $d(u, v) = \mathrm{diam}(G)$, then u and v are *antipodal vertices* of G.

For a connected graph G with diameter d, an *antipodal coloring* of a connected graph G is defined in [Chartrand et al. 2002a] as an assignment $c : V(G) \to \mathbb{N}$ of colors to the vertices of G such that

$$|c(u) - c(v)| + d(u, v) \geq d,$$

MSC2000: primary 05C12, 05C45; secondary 05C78, 05C15.

Keywords: Hamiltonian labeling, detour distance.

Willem Renzema's research was supported in part by Lee Honors College Dean's Summer Research Award Program at Western Michigan University.

for every two distinct vertices u and v of G. In the case of paths of order $n \geq 2$, this gives

$$|c(u) - c(v)| + d(u, v) \geq n - 1.$$

Antipodal colorings of paths gave rise to the more general Hamiltonian colorings of graphs defined in terms of another distance parameter.

The *detour distance* $D(u, v)$ between two vertices u and v in a connected graph G is the length of a longest path between these two vertices. A $u-v$ path of length $D(u, v)$ is a $u - v$ *detour*. Thus if G is a connected graph of order n, then

$$d(u, v) \leq D(u, v) \leq n - 1,$$

for every two vertices u and v in G, and

$$D(u, v) = n - 1,$$

if and only if G contains a Hamiltonian $u-v$ path. Furthermore $d(u, v) = D(u, v)$ for every two vertices u and v in G if and only if G is a tree. As with standard distance, the detour distance is a metric on the vertex set of a connected graph.

A *Hamiltonian coloring* of a connected graph G of order n is a coloring

$$c : V(G) \to \mathbb{N}$$

of G such that

$$|c(u) - c(v)| + D(u, v) \geq n - 1,$$

for every two distinct vertices u and v of G. Consequently, if u and v are distinct vertices such that $|c(u) - c(v)| = k$ for some Hamiltonian coloring c of G, then there is a $u - v$ path in G missing at most k vertices of G. The *value* $\mathrm{hc}(c)$ of a Hamiltonian coloring c of G is the maximum color assigned to a vertex of G. The *Hamiltonian chromatic number* of G is the minimum value of Hamiltonian colorings of G. Hamiltonian colorings of graphs have been studied in [Chartrand et al. 2002b; 2005a; 2005b; Nebeský 2003; 2006].

For a connected graph G with diameter d, a *radio labeling* of G is defined in [Chartrand et al. 2001] as an assignment $c : V(G) \to \mathbb{N}$ of labels to the vertices of G such that

$$|c(u) - c(v)| + d(u, v) \geq d + 1,$$

for every two distinct vertices u and v of G. Thus for a radio labeling of a graph, colors assigned to adjacent vertices of G must differ by at least d, colors assigned to two vertices at distance 2 must differ by at least $d - 1$, and so on, up to two vertices at distance d (that is, antipodal vertices), whose colors are only required to differ. The *value* $\mathrm{rn}(c)$ of a radio labeling c of G is the maximum color assigned to a vertex of G. The *radio number* of G is the minimum value of a radio labeling of G. In the case of paths of order $n \geq 2$, this gives

$$|c(u) - c(v)| + d(u, v) \geq n.$$

In a similar manner, radio labelings of paths and detour distance in graphs give rise to a related labeling, which we introduce in this work.

A *Hamiltonian labeling* of a connected graph G of order n is an assignment $c : V(G) \to \mathbb{N}$ of labels to the vertices of G such that

$$|c(u) - c(v)| + D(u, v) \geq n,$$

for every two distinct vertices u and v of G. Therefore, in a Hamiltonian labeling of G, every two vertices are assigned distinct labels and two vertices u and v can be assigned consecutive labels in G only if G contains a Hamiltonian $u - v$ path. We can assume that every Hamiltonian labeling of a graph uses the integer 1 as one of its labels. The *value* $\mathrm{hn}(c)$ of a Hamiltonian labeling c of G is the maximum label assigned to a vertex of G by c, that is, $\mathrm{hn}(c) = \max\{c(v) : v \in V(G)\}$. The *Hamiltonian labeling number* $\mathrm{hn}(G)$ of G is the minimum value of Hamiltonian labelings of G, that is, $\mathrm{hn}(G) = \min\{\mathrm{hn}(c)\}$, where the minimum is taken over all Hamiltonian labelings c of G. A Hamiltonian labeling c of G with value $\mathrm{hn}(c) = \mathrm{hn}(G)$ is called a *minimum Hamiltonian labeling* of G. Therefore,

$$\mathrm{hn}(G) \geq n. \tag{1}$$

for every connected graph G of order n.

To illustrate these concepts, we consider the Petersen graph P. It is known that $\chi(P) = \mathrm{hc}(P) = 3$. In fact, it is observed in [Chartrand et al. 2005a] that every proper coloring of P is also a Hamiltonian coloring. On the other hand, since the order of P is 10, it follows that $\mathrm{hn}(P) \geq 10$. Observe that $D(u, v) = 8$ if $uv \in E(G)$ and $D(u, v) = 9$ if $uv \notin E(G)$. Thus if c is a Hamiltonian labeling of P, then $|c(u) - c(v)| \geq 2$ if $uv \in E(G)$ and $|c(u) - c(v)| \geq 1$ if $uv \notin E(G)$. Therefore, the labeling shown in Figure 1 is a Hamiltonian labeling and so $\mathrm{hn}(P) = 10$.

2. Bounds for Hamiltonian labeling numbers of graphs

It is convenient to introduce some notation. For a Hamiltonian labeling c of a graph G, an ordering $u_1, u_2, \ldots, u_n$ of the vertices of G is called the *c-ordering* of G if

$$1 = c(u_1) < c(u_2) < \ldots < c(u_n) = \mathrm{hn}(c).$$

We refer to [Chartrand and Zhang 2008] for graph theory notation and terminology not described in this paper. In order to establish a relationship between the

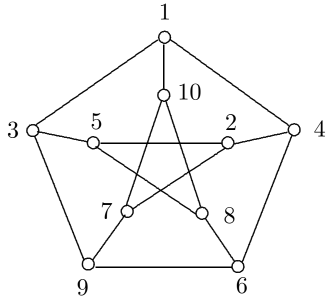


Figure 1. A Hamiltonian labeling of the Petersen graph.

Hamiltonian chromatic number and Hamiltonian labeling number of a connected graph, we first present a lemma.

Lemma 2.1. *Every connected graph of order $n \geq 3$ with Hamiltonian labeling number n is 2-connected.*

Proof. Assume, to the contrary, that there exists a connected graph G of order $n \geq 3$ with $\mathrm{hn}(G) = n$ such that G is not 2-connected. Then G contains a cut-vertex v. Let c be a minimum Hamiltonian labeling of G and let $v_1, v_2, \ldots, v_n$ be the c-ordering of the vertices of G, where then $1 = c(v_1) < c(v_2) < \ldots < c(v_n) = n$. Thus $c(v_i) = i$ for $1 \leq i \leq n$. Let $u \in V(G)$ such that u and v are consecutive in the c-ordering. Thus $\{u, v\} = \{v_j, v_{j+1}\}$ for some integer j with $1 \leq j \leq n-1$. Hence $D(v_j, v_{j+1}) \leq n-2$. However then,

$$|c(v_j) - c(v_{j+1})| + D(v_j, v_{j+1}) \leq n-1,$$

which contradicts the fact that c is a Hamiltonian labeling of G. $\square$

The corollary below now follows immediately.

Corollary 2.2. *No connected graph of order $n \geq 3$ with Hamiltonian labeling number n contains a bridge.*

While $\mathrm{hc}(K_1) = \mathrm{hn}(K_1) = 1$ and $\mathrm{hc}(K_2) = 1$ and $\mathrm{hn}(K_2) = 2$, $\mathrm{hc}(G)$ and $\mathrm{hn}(G)$ must differ by at least 2 for every connected graph G of order 3 or more. In fact, the following result provides upper and lower bounds for the Hamiltonian labeling number of a connected graph in terms of its order and Hamiltonian chromatic number.

Theorem 2.3. *For every connected graph G of order $n \geq 3$,*

$$\mathrm{hc}(G) + 2 \leq \mathrm{hn}(G) \leq \mathrm{hc}(G) + (n-1).$$

Proof. We first show that $\mathrm{hn}(G) \geq \mathrm{hc}(G)+2$. Let c be a minimum Hamiltonian labeling of G and let $v_1, v_2, \ldots, v_n$ be the c-ordering of the vertices of G, where then $1 = c(v_1) < c(v_2) < \ldots < c(v_n) = \mathrm{hn}(c)$. Define a coloring c^* of G by

$$c^*(v_i) = \begin{cases} 1 & \text{if } i = 1, \\ c(v_i) - 1 & \text{if } 2 \leq i \leq n-1, \\ c(v_i) - 2 & \text{if } i = n. \end{cases}$$

We show that c^* is a Hamiltonian coloring of G. Let $v_i, v_j \in V(G)$, where

$$1 \leq i < j \leq n.$$

We consider two cases.

Case 1. $i = 1$. Suppose first that $2 \leq j \leq n-2$. Then

$$|c^*(v_j) - c^*(v_1)| + D(v_j, v_1) = c(v_j) - c(v_1) - 1 + D(v_j, v_1) \geq n-1.$$

Next suppose that $j = n$. Then

$$\begin{aligned} |c^*(v_n) - c^*(v_1)| + D(v_n, v_1) &= c(v_n) - c(v_1) - 2 + D(v_n, v_1) \\ &= c(v_n) - 3 + D(v_n, v_1). \end{aligned}$$

If $c(v_n) \geq n+1$, then $c(v_n) - 3 + D(v_n, v_1) \geq n-1$. If $c(v_n) = n$, then $v_1 v_n$ is not a bridge by Corollary 2.2 and so $D(v_n, v_1) \geq 2$. Thus $c(v_n) - 3 + D(v_n, v_1) \geq n-1$.

Case 2. $i \geq 2$. In this case,

$$|c^*(v_j) - c^*(v_i)| + D(v_j, v_i) = \begin{cases} c(v_j) - c(v_i) + D(v_j, v_i), & \text{if } j \leq n-1, \\ c(v_j) - c(v_i) - 1 + D(v_j, v_i), & \text{if } j = n, \end{cases} \tag{2}$$

which is greater than or equal to $c(v_j) - c(v_i) - 1 + D(v_j, v_i) \geq n-1$. Thus c^* is a Hamiltonian coloring of G, as claimed. Therefore,

$$\mathrm{hc}(G) \leq \mathrm{hc}(c^*) = \mathrm{hn}(c) - 2 = \mathrm{hn}(G) - 2,$$

and so $\mathrm{hn}(G) \geq \mathrm{hc}(G) + 2$.

Next, we show that $\mathrm{hn}(G) \leq \mathrm{hc}(G) + (n-1)$. Let c' be a Hamiltonian coloring of G such that $\mathrm{hc}(c') = \mathrm{hc}(G)$. We may assume that $V(G) = \{v_1, v_2, \ldots, v_n\}$ such that

$$1 = c'(v_1) \leq c'(v_2) \leq \ldots \leq c'(v_n) = \mathrm{hc}(c').$$

Define a labeling c'' of G by $c''(v_i) = c'(v_i) + (i-1)$ for $1 \leq i \leq n$. Let v_j and v_k be two distinct vertices of G. Then

$$\begin{aligned} |c''(v_j) - c''(v_k)| + D(v_j, v_k) &= |c'(v_j) - c'(v_k)| + |j-k| + D(v_j, v_k) \\ &\geq (n-1) + |j-k| \geq n, \end{aligned}$$

and so c'' is a Hamiltonian labeling of G. Since $\mathrm{hn}(c'') = \mathrm{hc}(c) + (n-1)$, it follows that $\mathrm{hn}(G) \le \mathrm{hc}(G) + (n-1)$. □

While the upper and lower bounds in Theorem 2.3 are sharp (as we will see later), both inequalities in Theorem 2.3 can be strict. For example, consider the Petersen graph P of order $n = 10$ and $\mathrm{hn}(P) = 10$. Thus

$$5 = \mathrm{hc}(P) + 2 < \mathrm{hn}(P) < \mathrm{hc}(P) + (n-1) = 12.$$

In fact, more can be said. The following result was established in [Chartrand et al. 2005a].

Theorem 2.4 [Chartrand et al. 2005a]. *If G is a Hamiltonian graph of order $n \ge 3$, then $\mathrm{hc}(G) \le n-2$. Furthermore, for each pair k, n of integers with $1 \le k \le n-2$, there is a Hamiltonian graph of order n with Hamiltonian chromatic number k.*

On the other hand, every Hamiltonian graph of order n has Hamiltonian labeling number n, as we show next.

Proposition 2.5. *If G is a Hamiltonian graph of order $n \ge 3$, then $\mathrm{hn}(G) = n$.*

Proof. Let $C : v_1, v_2, \ldots, v_{n+1} = v_1$ be a Hamiltonian cycle of G. Define the labeling c of G by $c(v_i) = i$ for $1 \le i \le n$. Let i, j be two integers with $1 \le i < j \le n$. If $j - i \le n/2$, then $D(v_i, v_j) \ge n - (j-i)$; while if $j - i > n/2$, then $D(v_i, v_j) \ge j - i$. In either case, $|c(v_i) - c(v_j)| + D(v_i, v_j) \ge n$. Thus c is a Hamiltonian labeling and so $\mathrm{hn}(G) = n$ by Equation (1). □

The converse of Proposition 2.5 is not true. For example, it is well known that the Petersen graph P is a nonHamiltonian graph of order 10 but $\mathrm{hn}(P) = 10$. Whether there exists a connected graph G of order $n \ge 3$ with $\mathrm{hn}(G) = n$ that is neither a Hamiltonian graph nor the Petersen graph is not known. The following realization result is a consequence of Theorem 2.4 and Proposition 2.5.

Corollary 2.6. *For each pair k, n of integers with $2 \le k \le n-1$, there exists a Hamiltonian graph G of order n such that $\mathrm{hn}(G) = \mathrm{hc}(G) + k$.*

In the remainder of this section, we consider the complete bipartite graphs $K_{r,s}$ of order $n = r + s \ge 3$, where $1 \le r \le s$. The Hamiltonian chromatic number of a complete bipartite graph has been determined in [Chartrand et al. 2005a]. For positive integers r and s with $r \le s$ and $r + s \ge 3$,

$$\mathrm{hc}(K_{r,s}) = \begin{cases} r & \text{if } r = s, \\ (s-1)^2 + 1 & \text{if } 1 = r < s, \\ (s-1)^2 - (r-1)^2 & \text{if } 2 \le r < s. \end{cases} \tag{3}$$

If $r \ge 2$, then $K_{r,r}$ is Hamiltonian and so $\mathrm{hn}(K_{r,r}) = n = 2r$ by Proposition 2.5. Thus, we may assume that $r < s$, beginning with $r = 1$.

Theorem 2.7. *For each integer $n \geq 3$,*

$$\mathrm{hn}(K_{1,n-1}) = n + (n-2)^2.$$

Proof. Let $G = K_{1,n-1}$ with vertex set $\{v, v_1, v_2, \ldots, v_{n-1}\}$, where v is the central vertex of G. By Equation (3) and Theorem 2.3, it suffices to show that

$$\mathrm{hn}(G) \geq n + (n-2)^2.$$

Let c be a minimum Hamiltonian labeling of G. Since no two vertices of G can be labeled the same, we may assume that

$$c(v_1) < c(v_2) < \ldots < c(v_{n-1}).$$

We consider three cases.

Case 1. $c(v) = 1$. Since $D(v_1, v) = 1$ and $D(v_i, v_{i+1}) = 2$ for $1 \leq i \leq n-2$, it follows that $c(v_1) \geq n$ and

$$c(v_{i+1}) \geq c(v_i) + (n-2) \geq c(v_1) + i(n-2) \geq n + i(n-2)$$

for all $1 \leq i \leq n-2$. This implies that

$$c(v_{n-1}) \geq n + (n-2)(n-2) = n + (n-2)^2.$$

Therefore, $\mathrm{hn}(G) = \mathrm{hn}(c) \geq n + (n-2)^2$.

Case 2. $c(v) = \mathrm{hn}(c)$. Then $1 = c(v_1) < c(v_2) < \ldots < c(v_{n-1}) < c(v)$. For each i with $2 \leq i \leq n-1$, it follows that

$$c(v_i) \geq c(v_1) + (i-1)(n-2) = 1 + (i-1)(n-2).$$

In particular, $c(v_{n-1}) \geq 1 + (n-2)^2$. Thus

$$c(v) \geq c(v_{n-1}) + n - 1 = n + (n-2)^2.$$

Therefore, $\mathrm{hn}(G) = \mathrm{hn}(c) \geq n + (n-2)^2$.

Case 3. $c(v_j) < c(v) < c(v_{j+1})$ for some j with $1 \leq j \leq n-2$. Thus

$$\begin{aligned} c(v_j) &\geq 1 + (j-1)(n-2), \\ c(v) &\geq c(v_j) + n - 1 \geq n + (j-1)(n-2), \\ c(v_{j+1}) &\geq c(v) + n - 1 \geq 2n - 1 + (j-1)(n-2). \end{aligned}$$

This implies that

$$\begin{aligned} c(v_{n-1}) &\geq (n-j-2)(n-2) + c(v_{j+1}) \\ &\geq (n-j-2)(n-2) + (2n-1) + (j-1)(n-2) \\ &= 2n - 1 + (n-3)(n-2) = n + 1 + (n-2)^2 > n + (n-2)^2. \end{aligned}$$

In each case, we have $\mathrm{hn}(G) \geq n + (n-2)^2$. □

We now consider $K_{r,s}$, where $2 \le r < s$, with partite sets V_1 and V_2 such that $|V_1| = r$ and $|V_2| = s$. Then

$$D(u,v) = \begin{cases} 2r-2 = n-s+r-2 & \text{if } u,v \in V_1, \\ 2r-1 = n-s+r-1 & \text{if } uv \in E(K_{r,s}), \\ 2r = n-s+r & \text{if } u,v \in V_2. \end{cases}$$

Consequently, if c is a Hamiltonian labeling of $K_{r,s}$ $(r < s)$, then

$$|c(u)-c(v)| \ge \begin{cases} s-r+2 & \text{if } u,v \in V_1, \\ s-r+1 & \text{if } uv \in E(K_{r,s}), \\ s-r & \text{if } u,v \in V_2. \end{cases}$$

Theorem 2.8. *For integers r and s with $2 \le r < s$,*

$$\mathrm{hn}(K_{r,s}) = (s-1)^2 - (r-1)^2 + s + r - 1.$$

Proof. By Equation (3) and Theorem 2.3, it suffices to show that

$$\mathrm{hn}(K_{r,s}) \ge (s-1)^2 - (r-1)^2 + s + r - 1.$$

Let $V_1 = \{u_1, u_2, \dots, u_r\}$ and $V_2 = \{v_1, v_2, \dots, v_s\}$ be the partite sets of $K_{r,s}$, and let c be a Hamiltonian labeling of $K_{r,s}$ and let $w_1, w_2, \dots, w_{r+s}$ be the c-ordering of the vertices of $K_{r,s}$. We define a V_1-*block* of $K_{r,s}$ to be a set

$$A = \{w_\alpha, w_{\alpha+1}, \dots, w_\beta\},$$

where $1 \le \alpha \le \beta \le r+s$, such that $A \subseteq V_1$, $w_{\alpha-1} \in V_2$ if $\alpha > 1$, and $w_{\beta+1} \in V_2$ if $\beta < r+s$. A V_2-*block* of $K_{r,s}$ is defined similarly. Let

$$A_1, A_2, \dots, A_p \quad (p \ge 1)$$

be the distinct V_1-blocks of $K_{r,s}$ such that if

$$w' \in A_i, \quad w'' \in A_j,$$

where $1 \le i < j \le p$, then $c(w') < c(w'')$. If $p \ge 2$, then $K_{r,s}$ contains V_2-blocks $B_1, B_2, \dots, B_{p-1}$ such that for each integer i $(1 \le i \le p-1)$ and for $w' \in A_i$, $w \in B_i$, $w'' \in A_{i+1}$, it follows that

$$c(w') < c(w) < c(w'').$$

The graph $K_{r,s}$ may contain up to two additional V_2-blocks, namely B_0 and B_p such that if $y \in B_0$ and $y' \in A_1$, then $c(y) < c(y')$; while if $z \in A_p$ and $z' \in B_p$, then $c(z) < c(z')$. If $p = 1$, then at least one of B_0 and B_1 must exist. Hence $K_{r,s}$ contains p V_1-blocks and $p-1+t$ V_2-blocks, where $t \in \{0, 1, 2\}$. Consequently, there are exactly

(a) $r-p$ distinct pairs $\{w_i, w_{i+1}\}$ of vertices, both of which belong to V_1;

(b) $2p-2+t$ distinct pairs $\{w_i, w_{i+1}\}$ of vertices, exactly one of which belongs to V_1;

(c) $s-(p-1+t)$ distinct pairs $\{w_i, w_{i+1}\}$ of vertices, both of which belong to V_2.

Since (1) the colors of every two vertices w_i and w_{i+1}, both of which belong to V_1, must differ by at least $s-r+2$, (2) the colors of every two vertices w_i and w_{i+1}, exactly one of which belongs to V_1, must differ by at least $s-r+1$, and (3) the colors of every two vertices w_i and w_{i+1}, both of which belong to V_2, must differ by at least $s-r$, it follows that

$$\begin{aligned} c(w_{r+s}) &\geq 1+(r-p)(s-r+2)+(2p-2+t)(s-r+1)+(s-(p-1+t))(s-r) \\ &= (s-1)^2-(r-1)^2+s+r-1+t. \end{aligned} \tag{4}$$

Since $\mathrm{hn}(K_{r,s}) \leq (s-1)^2-(r-1)^2+s+r-1$ and $t \geq 0$, it follows that $t=0$ and that $\mathrm{hn}(K_{r,s}) = (s-1)^2-(r-1)^2+s+r-1$. □

Combining Proposition 2.5 and Theorems 2.7 and 2.8, we obtain the following.

Corollary 2.9. *For integers r and s with $1 \leq r \leq s$,*

$$\mathrm{hn}(K_{r,s}) = \begin{cases} r+s & \text{if } r=s, \\ (s-1)^2+s+1 & \text{if } r=1 \text{ and } s \geq 2, \\ (s-1)^2-(r-1)^2+r+s-1 & \text{if } 2 \leq r < s. \end{cases}$$

3. Hamiltonian labeling numbers of subgraphs of coronas of Hamiltonian graphs

A common question in graph theory concerns how the value of a parameter is affected by making a small change in the graph. If G is a Hamiltonian graph and u and v are two nonadjacent vertices of G, then $G+uv$ is also Hamiltonian and so $\mathrm{hn}(G) = \mathrm{hn}(G+uv)$. On the other hand, if we add a pendant edge to a Hamiltonian graph G producing a nonHamiltonian graph H, then the Hamiltonian labeling number of H can be significantly larger than that of G, as we show in this section. We begin with those graphs obtained from a cycle or a complete graph by adding a single pendant edge.

Theorem 3.1. *If G is the graph of order $n \geq 5$ obtained from C_{n-1} by adding a pendant edge, then* $\mathrm{hn}(G) = 2n-2$.

Proof. Let $C : v_1, v_2, \dots, v_{n-1}, v_1$ and let $v_{n-1}v_n$ be the pendant edge of G. We first show that $\mathrm{hn}(G) \leq 2n-2$. Define a labeling c_0 of G by

$$c_0(v_i) = \begin{cases} 2i & \text{if } 1 \leq i \leq n-1, \\ 1 & \text{if } i=n. \end{cases}$$

We show that c_0 is a Hamiltonian labeling. First let

$$v_i, v_j \in V(C),$$

where $1 \le i < j \le n-1$. If $j - i \ge \frac{n-1}{2}$, then $D(v_i, v_j) = j - i$ and so

$$\begin{aligned} |c_0(v_i) - c_0(v_j)| + D(v_i, v_j) &= |2i - 2j| + (j - i) = 3(j - i) \\ &\ge 3\Big(\frac{n-1}{2}\Big) = \frac{3n}{2} - \frac{3}{2} \ge n, \end{aligned}$$

since $n \ge 3$. If $j - i \le \frac{n-1}{2}$, then $D(v_i, v_j) = (n-1) - (j-i)$ and so

$$\begin{aligned} |c_0(v_i) - c_0(v_j)| + D(v_i, v_j) &= 2(j - i) + [(n-1) - (j-i)] \\ &= n - 1 + (j - i) \ge n. \end{aligned}$$

Next, we consider each pair v_i, v_n where $1 \le i \le n-1$. Since $D(v_i, v_n) \ge n - i$ and $|c_0(v_i) - c_0(v_n)| \ge 2i - 1$, it follows that

$$|c_0(v_i) - c_0(v_n)| + D(v_i, v_n) \ge n + i - 1 \ge n.$$

Therefore, c_0 is a Hamiltonian labeling, as claimed.

Next, we show that $\mathrm{hn}(G) \ge 2n - 2$. Let c be a minimum Hamiltonian labeling of G. First, we make some observations.

(a) For each pair i, j with $1 \le i \ne j \le n-1$, $D(v_i, v_j) \le n-2$ and so $|c(v_i) - c(v_j)| \ge 2$.

(b) For each i with $i \in \{1, n-2\}$, $D(v_n, v_i) = n - 1$ and so $|c(v_n) - c(v_i)| \ge 1$.

(c) For each i with $1 \le i \le n-1$ and $i \notin \{1, n-2\}$, $D(v_n, v_i) \le n-2$ and so $|c(v_n) - c(v_i)| \ge 2$.

Let $u_1, u_2, \ldots, u_n$ be the c-ordering of the vertices of G and let

$$X = \{c(u_{i+1}) - c(u_i) : 1 \le i \le n-1\}.$$

By observations (a)–(c), at most two terms in X are 1. If at most one term in X is 1, then $\mathrm{hn}(c) = c(u_n) \ge 1 + 1 + 2(n-2) = 2n - 2$. If at least one term in X is 3 or more, then $\mathrm{hn}(c) = c(u_n) \ge 1 + 1 + 1 + 3 + 2(n-4) = 2n - 2$. Thus we may assume that exactly two terms in X are 1 and the remaining terms in X are 2. Then $v_n = u_i$ for some i with $2 \le i \le n-1$ and $\{v_1, v_{n-2}\} = \{u_{i-1}, u_{i+1}\}$, where $c(u_i) - c(u_{i-1}) = c(u_{i+1}) - c(u_i) = 1$. This implies that $v_{n-1} = u_j$ for some j with $1 \le j \le n$ and $j \ne i$. If $2 \le j \le n-1$, then $\{u_{j-1}, u_{j+1}\} \ne \{v_1, v_{n-2}\}$; if $j = 1$, then $u_2 \notin \{v_1, v_{n-2}\}$, for otherwise

$$\begin{aligned} |c(v_{n-1}) - c(v_n)| + D(v_{n-1}, v_n) &\le |c(v_{n-1}) - c(u_2)| + |c(u_2) - c(v_n)| + 1 \\ &\le 2 + 1 + 1 = 4 < n, \end{aligned}$$

which is impossible; if $j = n$, then $u_{n-1} \notin \{v_1, v_{n-2}\}$, for otherwise

$$\begin{aligned}|c(v_{n-1}) - c(v_n)| + D(v_{n-1}, v_n) &\leq |c(v_{n-1}) - c(u_{n-1})| + |c(u_{n-1}) - c(v_n)| + 1 \\ &\leq 2 + 1 + 1 = 4 < n,\end{aligned}$$

again, which is impossible. Therefore, for each j with $1 \leq j \leq n$, there exists $k \in \{j-1, j+1\}$ such that $u_k \notin \{v_1, v_{n-2}\}$. Assume, without loss of generality, that $u_{j-1} \notin \{v_1, v_{n-2}\}$. Since $D(u_{j-1}, u_j) \leq n-3$, it follows that $c(u_j) - c(u_{j-1}) \geq 3$, which is impossible since each term in X is at most 2. Thus, $\mathrm{hn}(G) \geq 2n-2$. □

Theorem 3.2. *If G is the graph of order $n \geq 4$ obtained from K_{n-1} by adding a pendant edge, then* $\mathrm{hn}(G) = 2n - 3$.

Proof. Let $V(K_{n-1}) = \{v_1, v_2, \ldots, v_{n-1}\}$ and let G be obtained from K_{n-1} by adding the pendant edge $v_{n-1}v_n$. We first show that $\mathrm{hn}(G) \leq 2n - 3$. Define a labeling c_0 of G by

$$c_0(v) = \begin{cases} 2i - 1 & \text{if } v = v_i \text{ for } 1 \leq i \leq n-1, \\ 2 & \text{if } v = v_n. \end{cases}$$

For each pair i, j of integers with $1 \leq i \neq j \leq n-1$,

$$D(v_i, v_j) = n - 2 \quad \text{and} \quad |c_0(v_i) - c_0(v_j)| \geq 2.$$

For each i with $1 \leq i \leq n-2$,

$$D(v_n, v_i) = n - 1 \quad \text{and} \quad |c_0(v_n) - c_0(v_i)| \geq 1.$$

Furthermore, $D(v_n, v_{n-1}) = 1$ and

$$|c_0(v_n) - c_0(v_{n-1})| \geq (2n-3) - 2 = 2n - 5 \geq n - 1$$

for $n \geq 4$. In each case,

$$D(v_i, v_j) + |c_0(v_i) - c_0(v_j)| \geq n,$$

for all i, j with $1 \leq i \neq j \leq n$. Therefore, c_0 is a Hamiltonian labeling and so $\mathrm{hn}(G) \leq \mathrm{hn}(c_0) = c_0(v_{n-1}) = 2n - 3$.

Next, we show that $\mathrm{hn}(G) \geq 2n-3$. Let c be a minimum Hamiltonian labeling of G. Suppose that the vertices of K_{n-1} in G can be ordered as $u_1, u_2, \ldots, u_{n-1}$ such that $c(u_1) < c(u_2) < \ldots < c(u_{n-1})$. Since

$$D(u_i, u_j) = n - 2$$

for $1 \leq i < j \leq n-1$, it follows that

$$|c(u_i) - c(u_j)| = c(u_j) - c(u_i) \geq 2.$$

This implies that

$$\mathrm{hn}(c) \geq c(u_{n-1}) \geq 1 + 2(n-2) = 2n-3.$$

Therefore, $\mathrm{hn}(G) \geq 2n-3$. □

Let G be a connected graph containing an edge e that is not a bridge. Then $G-e$ is connected. For every two distinct vertices u and v in $G-e$, the length of a longest $u-v$ path in $G-e$ does not exceed the length of a longest $u-v$ path in G. Thus every Hamiltonian labeling of $G-e$ is a Hamiltonian labeling of G. This observation yields the following useful lemma.

Lemma 3.3. *If F is a connected subgraph of a connected graph G, then*

$$\mathrm{hn}(G) \leq \mathrm{hn}(F).$$

The following is a consequence of Theorems 3.1 and 3.2 and Lemma 3.3.

Corollary 3.4. *Let H be a Hamiltonian graph of order $n-1 \geq 3$. If G is a graph obtained from H by adding a pendant edge, then*

$$2n-3 \leq \mathrm{hn}(G) \leq 2n-2.$$

Proof. Let C be a Hamiltonian cycle in H. If $H = C_{n-1}$, then $\mathrm{hn}(G) = 2n-2$ by Theorem 3.1; while if $H = K_{n-1}$, then $\mathrm{hn}(G) = 2n-3$ by Theorem 3.2. Thus, we may assume that $H \neq C_{n-1}$ and $H \neq K_{n-1}$. Let F be the graph obtained from K_{n-1} by adding a pendant edge and F' be the graph obtained from C_{n-1} by adding a pendant edge. Then G can be obtained from F by deleting nonbridge edges and F' can be obtained from G by deleting nonbridge edges. It then follows by Lemma 3.3 that $\mathrm{hn}(F) \leq \mathrm{hn}(G) \leq \mathrm{hn}(F')$ and so $2n-3 \leq \mathrm{hn}(G) \leq 2n-2$. □

In fact, there exists a Hamiltonian graph H of order $n-1$ such that adding a pendant edge at a vertex x of H produces a graph G with $\mathrm{hn}(G) = 2n-3$ but adding a pendant edge at a different vertex y of H produces a graph F with $\mathrm{hn}(F) = 2n-2$. For example, let H be the Hamiltonian graph obtained from the cycle $C : v_1, v_2, \ldots, v_{n-1}, v_1$ of order $n-1 \geq 4$ by adding the edge v_1v_{n-2}. If G is formed from H by adding a pendant edge at v_{n-1}, then $\mathrm{hn}(G) = 2n-3$; while if F is formed from H by adding the pendant edge v_1, then $\mathrm{hn}(F) = 2n-2$.

In order to study graphs obtained from a Hamiltonian graph by adding pendant edges, we first establish some additional definitions and notation. For a graph F, the *corona* $\mathrm{cor}(F)$ of F is that graph obtained from F by adding exactly one pendant edge at each vertex of F. For a connected graph G, the *core* $C(G)$ of G is obtained from G by successively deleting vertices of degree 1 until none remain. Thus, if G is a tree, then its core is K_1; while if G is not a tree, then the core of G is the induced subgraph F of maximum order with $\delta(F) \geq 2$. For each integer $k \geq 3$, let $\mathcal{H}_k$ be the set of nonHamiltonian graphs that can be obtained from a

Hamiltonian graph of order k by adding pendant edges to this graph in such a way that at most one pendant edge is added to each vertex of the graph. Thus if $G \in \mathcal{H}_k$, then there is a Hamiltonian graph H of order k such that G is a connected subgraph of $\mathrm{cor}(H)$ whose core is H. We now establish lower and upper bounds for the Hamiltonian labeling number of a graph in $\mathcal{H}_k$ in terms of the integer k and the order of the graph, beginning with a lower bound.

Theorem 3.5. *Let $G \in \mathcal{H}_k$ be a graph of order n and $k+1 \le n \le 2k$. Then*

$$\mathrm{hn}\,(G) \ge (n-1)(n-k) + (2k-n).$$

Proof. Suppose that H is a Hamiltonian graph of order $k \ge 3$ and that $H \cong C(G)$. If $H \ncong K_k$, then G can be obtained from some graph $F \in \mathcal{H}_k$ by deleting nonbridge edges from F, where $C(F) \cong K_k$, and $V(G-H) = V(F-K_k)$. That is, G and F possess the same end-vertices. It then follows by Lemma 3.3 that

$$\mathrm{hn}\,(F) \le \mathrm{hn}(G).$$

Therefore, it suffices to show that

$$\mathrm{hn}\,(F) \ge (n-1)(n-k) + (2k-n).$$

Let $V(F) = U \cup W$, where $U = V(K_k)$ and $W = V(F) - U$. First we make some observations:

(a) If $x, y \in U$, then $D(x, y) = k-1$.

(b) If $x, y \in W$, then $D(x, y) = k+1$.

(c) If $x \in U$ and $y \in W$, then $D(x, y) = 1$ if $xy \in E(F)$ and $D(x, y) = k$ otherwise.

Let c be a minimum Hamiltonian labeling of F and let $v_1, v_2, \ldots, v_n$ be the c-ordering of the vertices of F. We define the four subsets S_u, S_w, $S_{u,w}$, and $S_{w,u}$ of $V(F)$ as follows:

$$\begin{aligned} S_u &= \{v_i : v_{i-1}, v_i \in U \text{ for } 2 \le i \le n\}, \\ S_w &= \{v_i : v_{i-1}, v_i \in W \text{ for } 2 \le i \le n\}, \\ S_{u,w} &= \{v_i : v_{i-1} \in U \text{ and } v_i \in W \text{ for } 2 \le i \le n\}, \\ S_{w,u} &= \{v_i : v_{i-1} \in W \text{ and } v_i \in U \text{ for } 2 \le i \le n\}. \end{aligned}$$

Let $|S_u| = n_u$, $|S_w| = n_w$, $|S_{u,w}| = n_{u,w}$, $|S_{w,u}| = n_{w,u}$. Since

$$S_u \cup S_w \cup S_{u,w} \cup S_{w,u} = V(F) - \{v_1\},$$

it follows that

$$n_u + n_w + n_{u,w} + n_{w,u} = n - 1. \tag{5}$$

For each integer i with $2 \le i \le n$,

(A) if $v_i \in S_u$, then $c(v_i) - c(v_{i-1}) \ge n-k+1$ by (a);

(B) if $v_i \in S_w$, then $c(v_i) - c(v_{i-1}) \ge n-k-1$ by (b);

(C) if $v_i \in S_u \cup S_w$, then either $c(v_i) - c(v_{i-1}) \ge n-1$ or $c(v_i) - c(v_{i-1}) \ge n-k$ by (iii), and so $c(v_i) - c(v_{i-1}) \ge n-k$ in this case.

It then follows by (A)–(C) and (5) that

$$\begin{aligned}\mathrm{hn}(c) = c(v_n) &\ge 1 + n_u(n-k+1) + n_w(n-k-1) + (n_{u,w} + n_{w,u})(n-k)\\ &= 1 + (n_u + n_w + n_{u,w} + n_{w,u})(n-k) + (n_u - n_w)\\ &= 1 + (n-1)(n-k) + (n_u - n_w).\end{aligned}$$

We claim that $n_u - n_w \ge 2k-n-1$. Since

$$S_u \cup S_{u,w} = \{v_i : v_{i-1} \in U \text{ for } 2 \le i \le n\},$$

it follows that

$$|S_u \cup S_{u,w}| = \begin{cases} |U|-1 & \text{if } v_n \in U,\\ |U| & \text{otherwise;}\end{cases}$$

and so

$$n_u + n_{u,w} = k \text{ or } n_u + n_{u,w} = k-1. \tag{6}$$

Since

$$\begin{aligned}S_w \cup S_{u,w} &= \{v_i : \quad v_i \in W \text{ for } 2 \le i \le n\}\\ &= \begin{cases} W - \{v_1\} & \text{if } v_1 \in W,\\ W & \text{otherwise,}\end{cases}\end{aligned}$$

it follows that

$$n_w + n_{u,w} = n-k \text{ or } n_w + n_{u,w} = n-k-1. \tag{7}$$

By Equations (6) and (7), we obtain

$$n_u - n_w = (n_u + n_{u,w}) - (n_w + n_{u,w}) \ge (k-1) - (n-k) = 2k-n-1,$$

as claimed. Therefore,

$$\mathrm{hn}(G) = \mathrm{hn}(c) \ge 1 + (n-1)(n-k) + (n_u - n_w) \ge (n-1)(n-k) + (2k-n).$$

This completes the proof. □

Theorem 3.6. *Let $G \in \mathscr{H}_k$ be a graph of order n and $k+2 \le n \le 2k$. Then*

$$\mathrm{hn}\,(G) \le 1 + n + (n-k-1)^2 + (k-2)(n-k+1).$$

Proof. Suppose that H is a Hamiltonian graph of order $k \geq 3$ and that $H \cong C(G)$. If $H \ncong C_k$, then C_k can be obtained from H by deleting edges. Thus there exists $F \in \mathcal{H}_k$ such that $C(F) \cong C_k$ and F can be obtained from G by deleting edges that are not bridges. It then follows by Lemma 3.3 that

$$\text{hn}\,(G) \leq \text{hn}(F).$$

Therefore, we may assume that $H \cong C_k : x_1, x_2, \ldots, x_k, x_1$. Now let

$$X = \{x_1, x_2, \ldots, x_k\} \quad \text{and} \quad Y = V(G) - X = \{y_1, y_2, \ldots, y_{n-k}\}$$

such that y_i is adjacent to x_{j_i}, for $1 \leq i \leq n-k$, and $1 = j_1 < j_2 < \ldots < j_{n-k} \leq k$. For each i with $1 \leq i \leq n-k$, let

$$g_i = j_{i+1} - j_i - 1, \tag{8}$$

where $j_{n-k+1} = j_1$; that is, g_i is the number of vertices of degree 2 between x_{j_i} and $x_{j_{i+1}}$ on C_k. Thus if $x_{j_i} y_i \in E(G)$, then $x_{j_i+g_i+1} y_{i+1} \in E(G)$, for $1 \leq i \leq n-k$, and

$$\sum_{i=1}^{n-k} g_i = 2k - n.$$

Now define the labeling c of G by

$$c(v) = \begin{cases} 1 & \text{if } v = x_k, \\ 1 + n - k & \text{if } v = y_1, \\ c(y_{i-1}) + (n-k-1) + g_{i-1} & \text{if } v = y_i \text{ and } 2 \leq i \leq n-k, \\ c(y_{n-k}) + n - k + g_{n-k} & \text{if } v = x_1, \\ c(x_{j-1}) + (n-k+1) & \text{if } v = x_j \text{ and } 2 \leq j \leq k-1. \end{cases} \tag{9}$$

Thus the c-ordering of the vertices of G is

$$x_k, y_1, y_2, \ldots, y_{n-k}, x_1, x_2, \ldots, x_{k-1},$$

and by Equation (9)

$$\begin{aligned} &c(x_k) = 1, \\ &c(y_i) = 1 + n - k + (i-1)(n-k-1) + \sum_{\ell=1}^{i-1} g_\ell \text{ for } 1 \leq i \leq n-k, \\ &c(x_1) = 1 + n + (n-k-1)^2, \\ &c(x_j) = 1 + n + (n-k-1)^2 + (j-1)(n-k+1) \text{ for } 2 \leq j \leq k-1. \end{aligned} \tag{10}$$

Therefore, the value of c is

$$\text{hn}(c) = c(x_{k-1}) = 1 + n + (n-k-1)^2 + (k-2)(n-k+1).$$

Thus it remains to show that c is a Hamiltonian labeling of G. First, we make some observations. Let $u, v \in V(G)$, where $u \neq v$.

(α) If $u = x_i$ and $v = x_j$ where $1 \leq i \neq j \leq k$, then $D(u,v) = \max\{|i-j|, k-|i-j|\}$.

(β) If $u = y_i$ and $v = y_j$ where $1 \leq i < j \leq n-k$, then

$$D(u,v) = 2 + \max\left\{ j - i + \sum_{\ell=i}^{j-1} g_\ell, k - \left(j - i + \sum_{\ell=i}^{j-1} g_\ell \right) \right\}.$$

(γ) If $u = x_i$, $v \in Y$, and $vx_j \in E(G)$ where $1 \leq i, j \leq k$ (possibly $i = j$), then $D(u,v) = 1$ if $i = j$ and $D(u,v) = 1 + \max\{|i-j|, k-|i-j|\}$ if $i \neq j$.

We show that

$$D(u,v) + |c(u) - c(v)| \geq n, \tag{11}$$

for every pair u, v of distinct vertices of G. We consider three cases.

Case 1. $u, v \in X$. Let $u = x_i$ and $v = x_j$, where $1 \leq i, j \leq k$. We may assume, without loss of generality, that $i < j$. If $j = k$, then

$$\begin{aligned} |c(x_i) - c(x_j)| &= c(x_i) - c(x_k) \\ &= [1 + n + (n-k-1)^2 + (i-1)(n-k+1)] - 1 \geq n, \end{aligned}$$

and so condition (11) is satisfied. Thus we may assume that $j \neq k$.

If $j - i = 1$, then $D(x_i, x_j) = k - 1$ and $|c(x_i) - c(x_j)| = n - k + 1$. Thus (11) holds in this case. If $j - i \geq \frac{k}{2}$, then

$$\begin{aligned} D(x_i, x_j) + |c(x_i) - c(x_j)| &= c(x_j) - c(x_i) + D(x_i, x_j) \\ &= (j-i)(n-k+1) + (j-i) = (j-i)(n-k+2) \\ &\geq \frac{k}{2}(n-k+2) = k\left(\frac{n-k}{2} + 1\right) \geq 2k \geq n. \end{aligned}$$

If $2 \leq j - i \leq \frac{k}{2}$, then

$$\begin{aligned} D(x_i, x_j) + |c(x_i) - c(x_j)| &= c(x_j) - c(x_i) + D(x_i, x_j) \\ &= (j-i)(n-k+1) + (k - (j-i)) \\ &= (j-i)(n-k) + k \\ &\geq 2(n-k) + k = 2n - k \geq n. \end{aligned}$$

Case 2. $u, v \in Y$. Let $u = y_i$ and $v = y_j$, where $1 \leq i, j \leq n - k$. We may assume, without loss of generality, that $i < j$. Then

$$|c(y_i) - c(y_j)| = c(y_j) - c(y_i) = (j-i)(n-k-1) + \sum_{\ell=i}^{j-1} g_\ell.$$

If $j - i + \sum_{\ell=i}^{j-1} g_\ell \geq \frac{k}{2}$, then

$$D(y_i, y_j) = 2 + j - i + \sum_{\ell=i}^{j-1} g_\ell$$

by (β), and so

$$\begin{aligned} D(y_i, y_j) + |c(y_i) - c(y_j)| &= (j-i)(n-k-1) + \left(\sum_{\ell=i}^{j-1} g_\ell\right) + 2 + j - i + \left(\sum_{\ell=i}^{j-1} g_\ell\right) \\ &\geq (j-i)(n-k-1) + 2 + (j-i) + [k - 2(j-i)] \\ &= (j-i)(n-k-2) + k + 2 \geq n. \end{aligned}$$

If $1 \leq j - i + \sum_{\ell=i}^{j-1} g_\ell \leq \frac{k}{2}$, then

$$D(y_i, y_j) = 2 + k - \left(j - i + \sum_{\ell=i}^{j-1} g_\ell\right)$$

by (β), and so

$$\begin{aligned} D(y_i, y_j) + |c(y_i) - c(y_j)| &= (j-i)(n-k-1) + \left(\sum_{\ell=i}^{j-1} g_\ell\right) + 2 + k - \left(j - i + \sum_{\ell=i}^{j-1} g_\ell\right) \\ &= (j-i)(n-k-2) + k + 2 \\ &\geq n - k - 2 + k + 2 = n. \end{aligned}$$

Case 3. *One of u and v is in X and the other is in Y, say $u \in X$ and $v \in Y$.* Let $u = x_i$ and $v = y_j$, where $1 \leq i \leq k$ and $1 \leq j \leq n-k$. We consider two subcases, according to whether $x_i y_j \in E(G)$ or $x_i y_j \notin E(G)$.

Subcase 3.1. $x_i y_j \in E(G)$. We proceed by induction to show that

$$c(x_i) - c(y_j) \geq n - 1$$

when $x_i y_j \in E(G)$. For $i = j = 1$,

$$\begin{aligned} |c(x_1) - c(y_1)| = c(x_1) - c(y_1) &= [1 + n + (n-k-1)^2] - (1 + n - k) \\ &= (n-k-1)^2 + k \geq n - 1 \text{ for } n \geq k + 2. \end{aligned}$$

Assume that $c(x_i) - c(y_j) \geq n - 1$. Since $x_{i+1+g_j} y_{j+1} \in E(G)$ by (8), we show that $c(x_{i+1+g_j}) - c(y_{j+1}) \geq n - 1$. Observe that

$$c(x_{i+1+g_j}) = c(x_i) + (g_j + 1)(n-k+1) \quad \text{and} \quad c(y_{j+1}) = c(y_j) + (n-k-1) + g_j.$$

It then follows by the induction hypothesis that

$$\begin{aligned} c(x_{i+1+g_j}) - c(y_{j+1}) &\geq n-1+(g_j+1)(n-k+1)-(n-k-1)-g_j \\ &= n+1+g_j(n-k) \geq n-1. \end{aligned}$$

Therefore if $x_i y_j \in E(G)$, then $|c(x_i)-c(y_j)| + D(x_i, y_j) \geq n-1+1 = n$. Thus condition (11) is satisfied.

Subcase 3.2. $x_i y_j \notin E(G)$. Then $i \neq j$. By (8), if $y_j x_m \in E(G)$, then

$$\sum_{\ell=1}^{j-1} g_\ell = m - j,$$

and

$$\begin{aligned} D(x_i, y_j) + |c(x_i)-c(y_j)| &= c(x_i)-c(y_j)+D(x_i,x_m)+1 \\ &= (n-k-1)^2+(i-j)(n-k)+i+j-(m-j)+k-1+D(x_i,x_m) \\ &= [(n-k-1)^2+(i-j)(n-k)+k+2j-1]+(i-m)+D(x_i,x_m). \end{aligned} \tag{12}$$

Now observe, if $i > j$, then $(i-j)(n-k)+k \geq n$; whereas if $1 \leq i < j \leq n-k$, then

$$\begin{aligned} &(n-k-1)^2+(i-j)(n-k)+k+2j-1 \\ &\quad = [(n-k)^2-2(n-k)]+i(n-k)-j(n-k-2)+k \\ &\quad \geq [(n-k)^2-2(n-k)]+i(n-k)-[(n-k)^2-2(n-k)]+k \geq n. \end{aligned}$$

Therefore, by Equation (12)

$$D(x_i, y_j) + |c(x_i)-c(y_j)| \geq n+(i-m)+D(x_i,x_m). \tag{13}$$

We then have three possible situations. If $i > m$, then $i-m > 0$ and so by condition (13), (11) is satisfied. If $m > i$ and $m-i \geq k/2$, then $D(x_i,x_m) = m-i$ and so by (13)

$$D(x_i, y_j) + |c(x_i)-c(y_j)| \geq n+(i-m)+(m-i) = n.$$

Finally, if $m > i$ and $m-i \leq k/2$, then $D(x_i,x_m) = k-(m-i)$ and so from (13)

$$\begin{aligned} D(x_i, y_j) + |c(x_i)-c(y_j)| &\geq n+(i-m)+[k-(m-i)] \\ &= n+k-2(m-i) \geq n+k-k = n. \end{aligned}$$

For each situation, condition (11) is satisfied. Therefore c is a Hamiltonian labeling of G. $\square$

We now present two corollaries of Theorems 3.5 and 3.6.

Corollary 3.7. *If G is a graph of order n that is the corona of a Hamiltonian graph*, *then*

$$\mathrm{hn}(G) = \binom{n}{2}.$$

Proof. Suppose that H is a Hamiltonian graph of order $k \geq 3$ and that $G = \mathrm{cor}(H)$. Then the order of G is $n = 2k$. We show that

$$\mathrm{hn}(G) = \binom{n}{2} = k(2k-1).$$

If $H \neq C_k$ and $H \neq K_k$, then G can be obtained from $\mathrm{cor}(K_k)$ by deleting nonbridge edges and $\mathrm{cor}(C_k)$ can be obtained from G by deleting edges that are not bridges. It then follows by Lemma 3.3 that

$$\mathrm{hn}(\mathrm{cor}(K_k)) \leq \mathrm{hn}(G) \leq \mathrm{hn}\big(\mathrm{cor}(C_k)\big).$$

Therefore, it suffices to show that

$$k(2k-1) \leq \mathrm{hn}\big(\mathrm{cor}(K_k)\big) \text{ and } \mathrm{hn}\big(\mathrm{cor}(C_k)\big) \leq k(2k-1).$$

From Theorems 3.5 and 3.6, we find that

$$\mathrm{hn}(\mathrm{cor}(K_k)) \geq (2k-1)(2k-k) + (2k-2k) = k(2k-1)$$

and

$$\begin{aligned} \mathrm{hn}(\mathrm{cor}(C_k)) &\leq 1 + 2k + (2k-k-1)^2 + (k-2)(2k-k+1) \\ &= 1 + 2k + k^2 - 2k + 1 + k^2 - k - 2 = k(2k-1). \end{aligned}$$

Therefore, $\mathrm{hn}(G) = k(2k-1)$. □

Corollary 3.8. *For each graph $G \in \mathcal{H}_k$*,

$$2k - 1 \leq \mathrm{hn}\,(G) \leq k(2k-1).$$

Proof. Let

$$f(x) = (x-1)(x-k) + (2k-x),$$

for $k+1 \leq x \leq 2k$ and let

$$g(x) = 1 + x + (x-k-1)^2 + (k-2)(x-k+1),$$

for $k+2 \leq x \leq 2k$. Let $G \in \mathcal{H}_k$ be a graph of order n where $k+1 \leq n \leq 2k$. Then by Corollary 3.4 and Theorems 3.5 and 3.6,

$$f(n) \leq \mathrm{hn}\,(G) \leq g(n).$$

Since each $f(x)$ and $g(x)$ is an increasing function in its domain, it follows that $f(x) \geq f(k+1) = 2k-1$ and $g(x) \leq g(2k) = k(2k-1)$, implying the desired result. □

Both lower and upper bound in Corollary 3.8 are sharp. For example, if $G' \in \mathscr{H}_k$ is a graph of order $k+1$ whose core is K_k, then $\mathrm{hn}(G') = 2n-3 = 2k-1$ by Theorem 3.2; while if $G'' \in \mathscr{H}_k$ is a graph of order $2k$ whose core is K_k, then

$$\mathrm{hn}(G'') = \binom{n}{2} = k(2k-1)$$

by Corollary 3.7.

Acknowledgements

We are grateful to the referee whose valuable suggestions resulted in an improved paper.

References

[Chartrand and Zhang 2008] G. Chartrand and P. Zhang, *Chromatic graph theory*, Chapman & Hall/CRC, Boca Raton, FL, 2008.

[Chartrand et al. 2001] G. Chartrand, D. Erwin, P. Zhang, and F. Harary, "Radio labelings of graphs", *Bull. Inst. Combin. Appl.* **33** (2001), 77–85. MR 2003d:05185 Zbl 0989.05102

[Chartrand et al. 2002a] G. Chartrand, D. Erwin, and P. Zhang, "Radio antipodal colorings of graphs", *Math. Bohem.* **127**:1 (2002), 57–69. MR 2003b:05065 Zbl 0995.05056

[Chartrand et al. 2002b] G. Chartrand, L. Nebeský, and P. Zhang, "Bounds for the Hamiltonian chromatic number of a graph", pp. 113–125 in *Proceedings of the Thirty-third Southeastern International Conference on Combinatorics, Graph Theory and Computing (Boca Raton, FL, 2002)*, vol. 157, 2002. MR 2004e:05062 Zbl 1029.05059

[Chartrand et al. 2005a] G. Chartrand, L. Nebeský, and P. Zhang, "Hamiltonian colorings of graphs", *Discrete Appl. Math.* **146**:3 (2005), 257–272. MR 2005i:05061 Zbl 1056.05054

[Chartrand et al. 2005b] G. Chartrand, L. Nebesky, and P. Zhang, "On Hamiltonian colorings of graphs", *Discrete Math.* **290**:2-3 (2005), 133–143. MR 2006a:05058 Zbl 1059.05046

[Nebeský 2003] L. Nebeský, "Hamiltonian colorings of graphs with long cycles", *Math. Bohem.* **128**:3 (2003), 263–275. MR 2004h:05051 Zbl 1050.05055

[Nebeský 2006] L. Nebeský, "The Hamiltonian chromatic number of a connected graph without large Hamiltonian-connected subgraphs", *Czech. Math. J.* **56**:2 (2006), 317–338. MR 2007k:05125

Received: 2008-08-21 Revised: Accepted: 2008-11-15

Wille.a.renzema@wmich.edu *Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008, United States*

Ping.zhang@wmich.edu *Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008, United States*

INVOLVE 2:1(2009)

Some characterizations of type-3 slant helices in Minkowski space-time

Melih Turgut and Süha Yilmaz

(Communicated by Frank Morgan)

In this work, the concept of a slant helix is extended to Minkowski space-time. In an analogous way, we define type-3 slant helices whose trinormal lines make a constant angle with a fixed direction. Moreover, some characterizations of such curves are presented.

1. Introduction

Many important results in the theory of curves in E^3 were initiated by G. Monge; G. Darboux pioneered the moving frame idea. Thereafter, F. Frenet defined his moving frame and his special equations, which play an important role in mechanics and kinematics as well as in differential geometry (for more details see [Boyer 1968]). At the beginning of the twentieth century, Einstein's theory opened a door to new geometries such as Minkowski space-time, which is simultaneously the geometry of special relativity and the geometry induced on each fixed tangent space of an arbitrary Lorentzian manifold.

In the case of a differentiable curve, at each point a tetrad of mutually orthogonal unit vectors (called tangent, normal, binormal, and trinormal) was defined and constructed, and the rates of change of these vectors along the curve define the curvatures of the curve in the space E_1^4 [O'Neill 1983]. Helices (inclined curves) are a well-known concept in classical differential geometry [Millman and Parker 1977].

The notion of a slant helix is due to Izumiya and Takeuchi [2004], who defined a slant helix in E^3 as a curve $\varphi = \varphi(s)$ with nonvanishing first curvature if the principal lines of φ make a constant angle with a fixed direction. In the same space, spherical images, the tangent and the binormal indicatrix and some characterizations of such curves are presented in [Kula and Yayli 2005].

In recent years, the theory of degenerate submanifolds has been treated by researchers and some classical differential geometry topics have been extended to

MSC2000: 53C40, 53C50.

Keywords: Minkowski space-time, Frenet frame, slant helix, type-3 slant helix.

Lorentz manifolds. For instance, Duggal and Bejancu [1996] studied null curves in Lorentz manifolds and determined Frenet frame for lightlike curves. Ferrández et al. [2002] studied null generalized helices in Lorentz–Minkowski space. In the light of degenerate submanifold theory, Karadag and Karadag [2008] defined null slant helices and wrote some characterizations in E_1^3 and also proved that there does not exist a null slant helix in E_1^4. Some characterizations of the Cartan framed null generalized helix and the null slant helix having a nonnull axis in Lorentz–Minkowski space were given in [Erdogan and Yilmaz 2008].

In the literature, all works adopt the definition of a slant helix in [Izumiya and Takeuchi 2004] as one whose principal lines make a constant angle with a fixed direction. Some of them deal with null curves in Lorentz–Minkowski spaces. In this paper, we define a special slant helix whose *trinormal* lines make a constant angle with a fixed direction and call such curves *type-3 slant helices*. Additionally, we present some characterizations of curves in the space E_1^4.

2. Preliminaries

The basic elements of the theory of curves in the space E_1^4 are briefly presented here (a more complete elementary treatment can be found in [O'Neill 1983]).

Minkowski space-time E_1^4 is a Euclidean space E^4 provided with the standard flat metric given by

$$g = -dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2,$$

where (x_1, x_2, x_3, x_4) is a rectangular coordinate system in E_1^4.

Since g is an indefinite metric, recall that a vector $v \in E_1^4$ can have one of the three causal characterizations:

(i) it can be space-like if $g(v, v) > 0$ or $v = 0$;

(ii) it can be time-like if $g(v, v) < 0$;

(iii) and it can be null (light-like) if $g(v, v)$=0 and $v \neq 0$.

Similarly, an arbitrary curve $\alpha = \alpha(s)$ in E_1^4 can be locally space-like, time-like or null (light-like), if all of its velocity vectors $\alpha'(s)$ are respectively space-like, time-like, or null. Also, recall that the norm of a vector v is given by

$$\|v\| = \sqrt{|g(v, v)|}.$$

Therefore, v is a unit vector if $g(v, v) = \pm 1$. Next, vectors v, w in E_1^4 are said to be orthogonal if $g(v, w) = 0$. The velocity of the curve $\alpha(s)$ is given by $\|\alpha'(s)\|$. Let a and b be two space-like vectors in E_1^4, then there is a unique real number $0 \leq \delta \leq \pi$, called the angle between a and b, such that $g(a, b) = \|a\|.\|b\| \cos \delta$. Let $\vartheta = \vartheta(s)$ be a curve in E_1^4. If the tangent vector field of this curve forms a constant angle with a constant vector field U, then this curve is called a helix or an inclined curve. Recall that $\vartheta = \vartheta(s)$ is a slant helix if its principal lines make a constant angle with a fixed direction.

Denote by $\{T(s), N(s), B(s), E(s)\}$ the moving Frenet frame along the curve $\alpha(s)$ in the space E_1^4. Then T, N, B, E are, respectively, the tangent, the principal normal, the binormal, and the trinormal vector fields. A space-like or time-like curve $\alpha(s)$ is said to be parametrized by arclength function s, if $g(\alpha'(s), \alpha'(s)) = \pm 1$.

Let $\alpha(s)$ be a space-like curve in the space-time E_1^4, parametrized by arclength function s. Then, for the unit speed curve α with nonnull frame vectors, the following Frenet equations are given in [Walrave 1995]

$$\begin{bmatrix} T' \\ N' \\ B' \\ E' \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 & 0 \\ -\kappa & 0 & \tau & 0 \\ 0 & \tau & 0 & \sigma \\ 0 & 0 & \sigma & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \\ E \end{bmatrix},$$

where T, N, B, and E are mutually orthogonal vectors satisfying $g(B, B) = -1$ and $g(T, T) = g(N, N) = g(E, E) = 1$, where, κ, τ and σ are first, second and third curvature of the curve α, respectively.

3. Type–3 slant helix in Minkowski space-time

Definition 3.1. *A curve $\psi = \psi(s)$ is called a type-3 slant helix if the trinormal lines of ψ make a constant angle with a fixed direction in E_1^4.*

Recall that an arbitrary curve is called a W-curve if it has constant Frenet curvatures [İlarslan and Boyacıoğlu 2007].

Lemma 3.2. *A type-3 slant helix with curvatures $\kappa \neq 0$, $\tau \neq 0$, and $\sigma \neq 0$ cannot be a W-curve in E_1^4.*

Proof. Let $\psi = \psi(s)$ be a type-3 slant helix with curvatures $\kappa \neq 0$, $\tau \neq 0$, and $\sigma \neq 0$ and also be a W-curve in E_1^4. From the definition of type-3 slant helix, we have

$$g(E, U) = \cos\delta, \tag{1}$$

where U is a fixed direction (a constant space-like vector) and δ is a constant angle. Differentiating Equation (1) with respect to s, we easily get $\sigma g(B, U) = 0$, which yields B perpendicular to U. Therefore, we may express U as

$$U = u_1 T + u_2 N + u_3 E. \tag{2}$$

Differentiating Equation (2) with respect to s and considering the Frenet equations, we have the following system of equations

$$\frac{du_1}{ds} - u_2\kappa = 0, \qquad \frac{du_2}{ds} + u_1\kappa = 0, \qquad u_2\tau + u_3\sigma = 0, \qquad \frac{du_3}{ds} = 0. \tag{3}$$

From $(3)_4$ we easily obtain $u_3 = \cos\delta = \text{constant} \neq 0$. Substituting this into $(3)_3$, (owing to $\sigma = \text{constant}$ and $\tau = \text{constant}$) we have $u_2 = -\sigma/\tau \cos\delta = \text{constant}$. Using this in $(3)_2$, we get $u_1 = 0$. This result and Equation $(3)_1$ imply that $u_2 = 0$, which is a contradiction. Therefore, type-3 slant helix cannot be a W-curve in E_1^4. □

Theorem 3.3. *Let $\psi = \psi(s)$ be a space-like curve with curvatures $\kappa \neq 0$, $\tau \neq 0$ and $\sigma \neq 0$ in E_1^4. Then $\psi = \psi(s)$ is a type-3 slant helix if and only if*

$$\left(\frac{\sigma}{\tau}\right)^2 + \left[\frac{1}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\right]^2 = \textit{constant}. \tag{4}$$

Proof. Let $\psi = \psi(s)$ be a type-3 slant helix in E_1^4. Then the equations in (3) hold. Thus, we easily have $u_3 = \cos\delta = \text{constant} \neq 0$ and

$$u_2 = -\frac{\sigma}{\tau}\cos\delta. \tag{5}$$

If we consider $(3)_1$ and $(3)_2$, we obtain a second order differential equation with respect to u_2 as follows:

$$\frac{d}{ds}\left[\frac{1}{\kappa}\frac{du_2}{ds}\right] + u_2\kappa = 0. \tag{6}$$

Using an exchange variable $t = \int_0^s \kappa\, ds$ in Equation (6),

$$\frac{d^2u_2}{dt^2} + u_2 = 0 \tag{7}$$

is obtained. Solution of Equations (7) and (5) gives us

$$A\cos\int_0^s \kappa\, ds + B\sin\int_0^s \kappa\, ds = -\frac{\sigma}{\tau}\cos\delta, \tag{8}$$

where A and B are real numbers. Differentiating Equation (8) with respect to s, we obtain

$$-A\kappa\sin\int_0^s \kappa\, ds + B\kappa\cos\int_0^s \kappa ds = -\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\cos\delta. \tag{9}$$

In terms of (8) and (9), coefficients A and B can be calculated by the Cramer method. They are obtained as

$$\begin{aligned} A &= -\left(\frac{\sigma}{\tau}\cos\delta\right)\cos\int_0^s \kappa\, ds + \frac{\cos\delta}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\sin\int_0^s \kappa\, ds, \\ B &= -\frac{\cos\delta}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\cos\int_0^s \kappa ds - \left(\frac{\sigma}{\tau}\cos\delta\right)\sin\int_0^s \kappa\, ds. \end{aligned} \tag{10}$$

If we form A^2+B^2, we get

$$\left(\frac{\sigma}{\tau}\right)^2+\left[\frac{1}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\right]^2=\frac{A^2+B^2}{(\cos\delta)^2}=\text{constant}. \tag{11}$$

Conversely, let us consider a vector given by

$$U=\left\{\frac{\sigma}{\tau}T+\frac{1}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)N+E\right\}\cos\delta. \tag{12}$$

Differentiating vector U and considering the differential of (11), we get

$$\frac{dU}{ds}=0, \tag{13}$$

where δ is a constant angle. Equation (13) shows that U is a constant vector. And then considering a space-like curve $\psi=\psi(s)$ with nonvanishing curvatures, we have

$$g(E,U)=\cos\delta, \tag{14}$$

and it follows that $\psi=\psi(s)$ is a type-3 slant helix in E_1^4. □

Now, considering the differential of (11) and (12), we give the following result and remark.

Corollary 3.4. *Let $\psi=\psi(s)$ be a space-like curve with curvatures $\kappa\neq 0$, $\tau\neq 0$, and $\sigma\neq 0$. Then ψ is a type-3 slant helix in E_1^4 if and only if*

$$\frac{\kappa\sigma}{\tau}+\frac{d}{ds}\left[\frac{1}{\kappa}\frac{d}{ds}\left(\frac{\sigma}{\tau}\right)\right]=0. \tag{15}$$

Remark 3.5. *The fixed direction in the definition of type-3 slant helix can be taken as Equation (12).*

Let us solve Equation (15) with respect to σ/τ. Using the exchange variable $t=\int_0^s\kappa\,ds$ in (15), we get

$$\frac{d^2}{dt^2}\left(\frac{\sigma}{\tau}\right)+\left(\frac{\sigma}{\tau}\right)=0, \tag{16}$$

and so

$$\frac{\sigma}{\tau}=L_1\cos\int_0^s\kappa\,ds+L_2\sin\int_0^s\kappa\,ds, \tag{17}$$

where L_1 and L_2 are real numbers.

Corollary 3.6. *Let $\psi=\psi(s)$ be a space-like curve with curvatures $\kappa\neq 0$, $\tau\neq 0$, and $\sigma\neq 0$, then ψ is a type-3 slant helix in E_1^4 if and only if there is the following relation among the curvatures of ψ:*

$$\frac{\sigma}{\tau}=L_1\cos\int_0^s\kappa ds+L_2\sin\int_0^s\kappa\,ds. \tag{18}$$

Acknowledgments

The authors are grateful to the referee for his/her critical comments which improved the first version of the paper.

References

[Boyer 1968] C. B. Boyer, *A history of mathematics*, John Wiley & Sons Inc., New York, 1968. MR 38 #3105

[Duggal and Bejancu 1996] K. L. Duggal and A. Bejancu, *Lightlike submanifolds of semi-Riemannian manifolds and applications*, vol. 364, Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996. MR 97e:53121

[Erdogan and Yilmaz 2008] M. Erdogan and G. Yilmaz, "Null Generalized and Slant Helices in 4-Dimensional Lorentz-Minkowski Space", *Int. J. Contemp. Math. Sci.* **3**:23 (2008), 1113–1120.

[Ferrández et al. 2002] A. Ferrández, A. Giménez, and P. Lucas, "Null generalized helices in Lorentz-Minkowski spaces", *J. Phys. A* **35**:39 (2002), 8243–8251. MR 2003j:53026 Zbl 1052.53028

[İlarslan and Boyacıoğlu 2007] K. İlarslan and Ö. Boyacıoğlu, "Position vectors of a spacelike W-curve in Minkowski space E_1^3", *Bull. Korean Math. Soc.* **44**:3 (2007), 429–438. MR 2348656

[Izumiya and Takeuchi 2004] S. Izumiya and N. Takeuchi, "New special curves and developable surfaces", *Turkish J. Math.* **28**:2 (2004), 153–163. MR 2005e:53002 Zbl 1081.53003

[Karadag and Karadag 2008] H. Karadag and M. Karadag, "Null Generalized Slant Helices in 4-Dimensional Lorentzian Space", *Differ. Geom. Dyn. Syst.* **10** (2008), 178–185.

[Kula and Yayli 2005] L. Kula and Y. Yayli, "On slant helix and its spherical indicatrix", *Appl. Math. Comput.* **169**:1 (2005), 600–607. MR 2006k:53008 Zbl 1083.53006

[Millman and Parker 1977] R. S. Millman and G. D. Parker, *Elements of differential geometry*, Prentice-Hall Inc., Englewood Cliffs, N. J., 1977. MR 56 #1208 Zbl 0425.53001

[O'Neill 1983] B. O'Neill, *Semi-Riemannian geometry*, vol. 103, Pure and Applied Mathematics, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983. With applications to relativity. MR 85f:53002 Zbl 0531.53051

[Walrave 1995] J. Walrave, *Curves and Surfaces in Minkowski Space*, Ph.D. thesis, K. U. Leuven, Faculty of Science, 1995.

Received: 2008-03-17 Revised: 2008-06-14 Accepted: 2008-06-16

melih.turgut@gmail.com *Dokuz Eylül University, Buca Faculty of Education, Department of Science and Mathematics, 35160 Buca-Izmir, Turkey*

suha.yilmaz@yahoo.com *Dokuz Eylül University, Buca Faculty of Education, Department of Science and Mathematics, 35160 Buca-Izmir, Turkey*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the Involve website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *Involve* are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

involve

2009 Volume 2 No. 1