Quasigroups and Related Systems **14** (2006), 1 – 9

Loop algebras of loops whose derived subloop is central

Luiz G. X. de Barros

Abstract

The *isomorphism problem* for loops, that is, to know under which conditions the loop algebra isomorphism implies the loop isomorphism, is studied in the semisimple case for loops whose derived subloop is central. This is done by studying the structure of the semisimple loop algebra and by proving that it can be decomposed as a direct sum of an associative and commutative subalgebra with a nonassociative and non-commutative subalgebra.

1. Nomenclature and Introduction

A *loop* is a set L with a binary operation $\cdot$ which admits an identity element 1 and that the equations $X \cdot a = b$ and $a \cdot X = b$ have unique solutions for all a and b in L.

The unique solution of the equation $a \cdot b = (b \cdot a) \cdot X$ in L is called the *commutator* of the elements a and b, while the unique solution of the equation $(a \cdot b) \cdot c = (a \cdot (b \cdot c)) \cdot X$ is called the *associator* of the elements a,b and c. For a loop L, the subloop L' generated by all commutators and all associators is called the *derived subloop* of L. The quotient loop L/L' is an abelian group and L' is the smallest subloop of L with such property.

The *centre* of a loop L is the set $\mathcal{Z}(L)$ of the elements in L which commute with any element in L and which associate with any two others elements in L in any order of association.

2000 Mathematics Subject Classification: Primary 17A99; Secondary 20N05
Keywords: loop, loop algebra

In a loop, the solution of the equation $X \cdot a = 1$ is called the *left inverse* of the element a and it is denoted by a^{λ}, while the solution of the equation $a \cdot X = 1$ is called the *right inverse* of a and it is denoted by a^{ρ}.

Given an associative and commutative ring R and a loop L, the *loop ring* RL is the free R-module with basis L and multiplication defined distributively from the multiplication of L. For a field K, the *loop algebra* KL is defined in the same way.

For a normal subloop N of a loop L , the epimorphism $L \longrightarrow L/N$ extends to the algebra epimorphism $KL \longrightarrow K[L/N]$ whose kernel, denoted by $\Delta(L, N)$, is the ideal of KL generated by the set $\{n - 1 \mid n \in N\}$.

We shall denote by $[KL, KL]$ the left ideal of KL generated by all elements of the form $\alpha\beta - \beta\alpha$ with $\alpha, \beta \in KL$, and by $[KL, KL, KL]$ the left ideal of KL generated by all elements of the form $\alpha\beta \cdot \gamma - \alpha \cdot \beta\gamma$ with $\alpha, \beta, \gamma \in KL$.

The *isomorphism problem* for group rings, posed by Graham Higman in his 1940 thesis, asks if a group is *determined* by its group ring, that is, given a ring R and groups G and H, does the ring isomorphism $RG \cong RH$ imply the group isomorphism $G \cong H$?

G. Higman himself proved that the integral group ring of a finite abelian group determines the group. Later, A. Whitcomb extended this result to integral group ring of finite metabelian groups.

The isomorphism problem over fields was first considered in 1950 by S. Perlis and G.L. Walker who proved that if G is a finite abelian group then G is determined its rational group algebra $\mathbf{Q}G$. Solutions for many classes of group, many kinds of rings and fields are found during this time.

On the '80s, the problem was reposed for loops and expressive results for RA-loops were obtained. For details, Chapter XI of the book *"Alternative Loop Rings"*, Elsevier, (1996), by E.G. Goodaire, E. Jespers and C. Polcino Milies, is recommended.

During the '90s, the author has worked in extending these results to other classes of loops. In this work, some advances on the loops whose derived subloop is central are shown. The semisimple case, that is, when the characteristic of the field does not divide the order of the loop, is completely solved. This is made studying the structure of the loop algebra and proving that it can be decomposed as a direct sum of an associative and commutative subalgebra with a non-associative and non-commutative subalgebra.

2. Loop algebras of loops whose derived subloop is central

Here we focus our attention on the class $\mathcal{L}$ of the finite loops whose derived subloop is central. These loops appear very often in the theory of loops. Code loops [6], RA-loops [5], loops with a unique non-trivial commutator-associator element [8] are examples of such loops.

Some of the results in this section are extensions of those obtained for RA-loops in [1] by the author and in [3] by C. Polcino Milies and the author.

This first lemma is fundamental for the sequence and extends to loop algebras of loops in $\mathcal{L}$ a very known result of group algebras.

Lemma 2.1. *Let K be a field and L be a loop in $\mathcal{L}$ with derived subloop L'. Then*

$$[KL, KL] + [KL, KL, KL] = \Delta(L, L').$$

Proof. First, we observe that $[KL, KL]$ is generated by elements of the form $lm - ml$ with $l, m \in L$. Since $lm = ml \cdot (l, m)$ we have that

$$lm - ml = ml \cdot (l, m) - ml = ml \cdot ((l, m) - 1) \in \Delta(L, L') \ .$$

Also, $[KL, KL, KL]$ is generated by elements of the form $lm.n - l.mn$ with $l, m, n \in L$. Since, $lm.n = (l.mn) \cdot (l, m, n)$ we have that

$$lm.n - l.mn = (l.mn) \cdot (l, m, n) - l.mn = (l.mn) \cdot ((l, m, n) - 1) \in \Delta(L, L') \ .$$

Thus $[KL, KL] + [KL, KL, KL] \subset \Delta(L, L')$.

On the other hand, since $lm = ml \cdot (l, m)$ and (l, m) is central, we have that $(l, m) = (ml)^{\lambda} \cdot lm$ and $(l, m)^{-1} = (lm)^{\lambda} \cdot ml$.

Then

$$\begin{aligned} 1 - (l, m) &= 1 - (ml)^{\lambda} \cdot lm = (ml)^{\lambda} \cdot ml - (ml)^{\lambda} \cdot lm \\ &= (ml)^{\lambda} \cdot (ml - lm) \in [KL, KL] \end{aligned}$$

and

$$\begin{aligned} 1 - (l, m)^{-1} &= 1 - (lm)^{\lambda} \cdot ml = (lm)^{\lambda} \cdot lm - (lm)^{\lambda} \cdot ml \\ &= (lm)^{\lambda} \cdot (lm - ml) \in [KL, KL]. \end{aligned}$$

Also, since $lm.n = (l.mn) \cdot (l, m, n)$ and (l, m, n) is central, we have that $(l, m, n) = (l.mn)^{\lambda} \cdot (lm.n)$ and $(l, m, n)^{-1} = (lm.n)^{\lambda} \cdot (l.mn)$.

Then

$$1-(l,m,n) = 1-(l.mn)^{\lambda}\cdot(lm.n) = (l.mn)^{\lambda}\cdot(l.mn)-(l.mn)^{\lambda}\cdot(lm.n)$$
$$= (l.mn)^{\lambda}\cdot(l.mn-lm.n)\in[KL,KL,KL]$$

and

$$1-(l,m,n)^{-1}1-(lm.n)^{\lambda}\cdot(l.mn) = (lm.n)^{\lambda}\cdot(lm.n)-lm.n)^{\lambda}\cdot(l.mn)$$
$$= (lm.n)^{\lambda}\cdot(lm.n-l.mn)\in[KL,KL,KL].$$

An element $x\in L'$ is a product of commutators, associators and their inverses. The identity $1-cd=(1-c)+(1-d)-(1-c)\cdot(1-d)$ shows that $1-x$ can be separated in terms of the form $\alpha\cdot(1-c)$ where $\alpha\in KL$ and c is a commutator (or its inverse) or an associator (or its inverse). Each one of these terms of the form $\alpha\cdot(1-c)$ belongs to $[KL,KL]$ or $[KL,KL,KL]$. Thus $\Delta(L,L')\subset[KL,KL]+[KL,KL,KL]$. □

For a future use we recall a classical result about group algebras of finite abelian groups due to S. Perlis and G.L. Walker [10],

Theorem 2.2. (Theorem X.2.1, [7]) *Let G be a finite abelian group of order n and K be field whose characteristic does not divide n. Then*

$$KG\cong\oplus_{d|n}a_dK(\xi_d)\ ,$$

where $a_d=n_d/[K(\xi_d):K]$, n_d denotes the number of elements of order d in G and ξ_d denotes a primitive dth root of unity over K.

Using Lemma VI.1.2 of [7], we can prove

Lemma 2.3. *Let L be a loop in $\mathcal{L}$ and let K be any field whose characteristic does not divide $|L'|$, the order of L'. Define $\hat{L}'=\frac{1}{|L'|}\cdot\sum_{n\in L'}n$. Then*

i) $\hat{L}'$ *is a central idempotent in* KL,
ii) $KL\cdot\hat{L}'\cong K[L/L']$ *and* $KL\cdot(1-\hat{L}')\cong\Delta(L,L')$,
iii) $KL\cong KL\cdot\hat{L}'\oplus KL\cdot(1-\hat{L}')\cong K[L/L']\oplus\Delta(L,L')$.

The next result is valid for any field.

Proposition 2.4. *Let L_1 and L_2 be loops in $\mathcal{L}$ and K be any field. Suppose that $KL_1\cong KL_2$. Then $K[L_1/L_1']\cong K[L_2/L_2']$ and $\Delta(L_1,L_1')\cong\Delta(L_2,L_2')$.*

Proof. Let $KL_1 \longrightarrow K[L_1/L_1']$ be the natural epimorphism whose kernel is $\Delta(L_1, L_1')$.

Given an isomorphism $\Psi : KL_1 \longrightarrow KL_2$, we have

$$\Psi(\Delta(L_1, L_1')) = \Psi([KL_1, KL_1] + [KL, KL, KL]) =$$
$$= [KL_2, KL_2] + [KL_2, KL_2, KL_2] = \Delta(L_2, L_2').$$

This shows that Ψ also induces an isomorphism $\overline{\Psi}$ of the corresponding factor rings, so

$$K[L_1/L_1'] \cong KL_1/\Delta(L_1, L_1') \cong KL_2/\Delta(L_2, L_2') \cong K[L_2/L_2']. \qquad \square$$

As a consequence of both results, we obtain the next theorem which is an extension of Theorem 3.2 in [2] for RA-loops

Theorem 2.5. *Let L_1 and L_2 be loops in $\mathcal{L}$ and K be any field whose characteristic does not divide $|L_1'|$ and $|L_2'|$. Then $KL_1 \cong KL_2$ if and only if $K[L_1/L_1'] \cong K[L_2/L_2']$ and $\Delta(L_1, L_1') \cong \Delta(L_2, L_2')$.*

Using Theorem 2.2 we have

Corollary 2.6. *Let L_1 and L_2 be loops in $\mathcal{L}$ and let $\mathbf{Q}$ be the rational field. Then $\mathbf{Q}L_1 \cong \mathbf{Q}L_2$ if and only if $L_1/L_1' \cong L_2/L_2'$ and $\Delta(L_1, L_1') \cong \Delta(L_2, L_2')$.*

3. A subclass of $\mathcal{L}$

In this section we study the class $\mathcal{L}_1$ of the finite loops L such that $L/\mathcal{Z}(L) \cong C_2 \times C_2$. These loops appear as groups in the main papers about RA-loops as [5] and [9].

Proposition 3.1. *Let $L \in \mathcal{L}_1$. Then*

i) $L' \subset \mathcal{Z}(L)$, *that is,* $\mathcal{L}_1 \subset \mathcal{L}$,

ii) $L^2 \subset \mathcal{Z}(L)$,

iii) $L = < x, y, \mathcal{Z}(L) >$, *the subloop generated by x, y and $\mathcal{Z}(L)$, for all non-central elements $x, y \in L$ such that $x.\mathcal{Z}(L) \neq y.\mathcal{Z}(L)$.*

Proof.

i) It comes from the fact that $L/\mathcal{Z}(L)$ is an abelian group.

ii) It comes from the fact that the group $C_2 \times C_2$ has exponent 2.

iii) It comes from the fact that $L/\mathcal{Z}(L) \cong C_2 \times C_2$ can be generated by two non-central elements x and y such that $x.\mathcal{Z}(L) \neq y.\mathcal{Z}(L)$. $\square$

Proposition 3.2. *Let $L =< x, y, \mathcal{Z}(L) >$ a loop in $\mathcal{L}_1$. Write $\mathcal{Z}(L) \cong E \times A$, where E is an abelian 2-group and A is an abelian group of odd order. Then there exist $x_0, y_0 \in L$, with $x_0^2, y_0^2 \in E$ such that $L =< x_0, y_0, \mathcal{Z}(L) >$.*

Proof. Write $x^2 = x_1.x_2$ and $y^2 = y_1.y_2$ with $x_1, y_1 \in E$ and $x_2, y_2 \in A$. Let $x_0 = x^{o(x_2)}$ and $y_0 = y^{o(y_2)}$. It is easy to see that x_0 and y_0 have the desired properties. □

Theorem 3.3. *Let L be a loop in $\mathcal{L}_1$. Then $L = M \times A$, where M is a 2-loop in $\mathcal{L}_1$ and A is an abelian group of odd order.*

Proof. Write $\mathcal{Z}(L) \cong E \times A$, where E is an abelian 2-group and A is an abelian group of odd order. By Proposition 3.2 there exist $x_0, y_0 \in L$, with $x_0^2, y_0^2 \in E$ such that $L =< x_0, y_0, \mathcal{Z}(L) >$. Define $M =< x_0, y_0, E >$. Then $\mathcal{Z}(M) = E$ and $M/\mathcal{Z}(M) \cong C_2 \times C_2$. Also $L = M \times A$. □

Corollary 3.4. *Let L be a loop in $\mathcal{L}_1$. Write $L = M \times A$, where M is a 2-loop in $\mathcal{L}_1$ and A is an abelian group of odd order. Then $L' = M'$.*

Proof. Given the elements $l = x_0^a.y_0^b.z_l$, $m = x_0^c.y_0^d.z_m$ and $n = x_0^e.y_0^f.z_n$ in L with $x_o, y_0 \in M$ and $z_l, z_m, z_n \in \mathcal{Z}(L)$, observe that

$$(l, m) = (x_0^a.y_0^b.z_l \ , \ x_0^c.y_0^d.z_m) = (x_0^a.y_0^b \ , \ x_0^c.y_0^d) \in M'$$

and

$$(l, m, n) = (x_0^a.y_0^b.z_l \ , \ x_0^c.y_0^d.z_m \ , \ x_0^c.y_0^d.z_n) = (x_0^a.y_0^b \ , \ x_0^c.y_0^d \ , \ x_0^c.y_0^d) \in M'$$

since z_l , z_m and z_n are central. □

We say that elements a and b of a loop L are *conjugate* if $b = \theta(a)$ for some $\theta \in Inn(L)$, the *inner mapping group* of L. Conjugacy defines an equivalence relation on L. In a loop ring, a (finite) *class sum* is the sum of all the elements in a finite conjugacy class of L.

The next theorem is a classical result due to R. H. Bruck and it appears in [7] as Theorem III.1.3

Theorem 3.5. *Let L be a loop and R be a commutative and associative ring. The (finite) class sums of the loop ring RL form a R-basis for the centre of RL .*

Now we focus our attention to the class $\mathcal{L}_2$ of the loops L in $\mathcal{L}_1$ with a unique nonidentity commutator-associator element.

As the Corollary III.1.5 in [7], we can prove

Corollary 3.6. *Let L be a loop in $\mathcal{L}_2$ with a unique nonidentity commutator-associator element s and let R be a commutative and associative ring. Then the centre of the loop ring RL is spanned by the centre of L and those elements of RL of the form $l + sl$, $l \in L$.*

Also, as the Corollary VI.1.3 in [7], we can prove

Lemma 3.7. *Let L be a loop in $\mathcal{L}_2$ with a unique nonidentity commutator-associator element s and let K be a field whose characteristic does not divide the order of L. Then $\mathcal{Z}(\Delta(L, L')) \cong K[\mathcal{Z}(L)](1 - \hat{L}')$, where $\hat{L}' = \frac{1+s}{2}$.*

The next theorem extends Theorem 2.5 of [4] from RA-loops to the class $\mathcal{L}_2$.

Theorem 3.8. *Let L_1 and L_2 be loops in $\mathcal{L}_2$ and K be a field whose characteristic does not divide the order of either of these loops. For $1 = 1, 2$, write $L_i = M_i \times A_i$ where M_i is 2-loop in $\mathcal{L}_2$ and A_i is an abelian group of odd order. Then $KL_1 \cong KL_2$ if and only if $KM_1 \cong KM_2$ and $KA_1 \cong KA_2$.*

Proof. Suppose first that $KL_1 \cong KL_2$. By Proposition 2.4, we have $K[L_1/L_1'] \cong K[L_2/L_2']$; that is,

$$K[(M_1/M_1') \times A_1] \cong K[(M_2/M_2') \times A_2].$$

As these group algebras are commutative, using a result due to D. E. Cohen and which appears as Theorem X.2.5 in [7], we can conclude that

$$K[M_1/M_1'] \cong K[M_2/M_2'] \quad \textit{and} \quad KA_1 \cong KA_2.$$

In view of Theorem 2.5, in order to prove that $KM_1 \cong KM_2$ as well, it will suffice to show that $\Delta(M_1, M_1') \cong \Delta(M_2, M_2')$. By Theorem 2.5, $\Delta(L_1, L_1') \cong \Delta(L_2, L_2')$. Moreover, for $i = 1, 2$, denoting by $\hat{L}_i' = \frac{1}{|L_i'|} \cdot \sum_{n \in L_i'} n$ the central idempotent in KL_i,

$$\Delta(L_i, L_i') = KL_i(1 - \hat{L}_i') \cong (KM_i \otimes KA_i)(1 - \hat{L}_i') \cong \Delta(M_i, M_i') \otimes KA_i$$

since $\hat{L}_i' \in M_i$. Thus

$$\Delta(M_1, M_1') \otimes KA_1 \cong \Delta(L_1, L_1') \cong \Delta(L_2, L_2') \cong \Delta(M_2, M_2') \otimes KA_2.$$

Using Theorem 2.2, we have that $KA_1 \cong nK \oplus (\oplus_d m_d K(\xi_d)) \cong KA_2$, where ξ_d is a primitive root of unity of odd order d and d runs over the set of divisors of $|A_1|$ such that $K(\xi_d) \neq K$. For $i = 1, 2$,

$$\Delta(L_i, L_i') \cong n\Delta(M_i, M_i') \oplus (\oplus_d m_d(\Delta(M_i, M_i') \otimes K(\xi_d))).$$

By Lemma 3.7 , $\mathcal{Z}(\Delta(M_i, M_i')) = K[\mathcal{Z}(M_i)](1 - \hat{L}_i')$. Thus, using again Theorem 2.2, $\mathcal{Z}(\Delta(M_i, M_i')) \cong \oplus_j K(\xi_{a_j})$, where the ξ_{a_j} are primitive roots of unity of order 2^{a_j}. Consequently,

$$\mathcal{Z}(\Delta(M_i, M_i') \otimes K(\xi_d)) \cong \oplus_j K(\xi_{a_j}) \otimes K(\xi_d).$$

Since $(d, 2^{a_j}) = 1$, we have that

$$K(\xi_{a_j}) \otimes K(\xi_d) \cong K(\xi_{a_j})(\xi_d) = K(\xi_{a_j d}) \ ,$$

where $\xi_{a_j d}$ is a primitive root of unity of order $2^{a_j} d$. We claim that this field is never isomorphic to a field of the form $K(\xi_{a_i})$. In fact, assume that $K(\xi_{a_j d}) \cong K(\xi_{a_i})$. Then $K(\xi_d) \subset K(\xi_{a_i})$ and so

$$K(\xi_{a_i}) \otimes_K K(\xi_d) \cong K(\xi_{a_i d}) = K(\xi_{a_i})(\xi_d) = K(\xi_{a_i}).$$

However, as $K(\xi_d) \neq K$, the tensor product $K(\xi_{a_i}) \otimes_K K(\xi_d)$ has dimension at least two over the field $K(\xi_{a_i})$, a contradiction. Hence the centre of the algebra $\Delta(M_i, M_i')$ is a direct sum of fields which are all different from those appearing in the decomposition of the centre of $\Delta(M_i, M_i') \otimes K(\xi_d)$. Since

$$n\Delta(M_1, M_1') \oplus (\oplus_d m_d(\Delta(M_1, M_1') \otimes K(\xi_d))) \cong$$

$$\cong n\Delta(M_2, M_2') \oplus (\oplus_d m_d(\Delta(M_2, M_2') \otimes K(\xi_d)))$$

and because $\Delta(M_i, M_i')$ is a sum of algebras over fields of the form $K(\xi_{a_j})$ while $\mathcal{Z}(\Delta(M_i, M_i') \otimes K(\xi_d))$ contains no such direct summands, it follows that $n\Delta(M_1, M_1') \cong n\Delta(M_2, M_2')$. Hence $\Delta(M_1, M_1') \cong \Delta(M_2, M_2')$, as desired.

The converse is straightforward. □

Corollary 3.9. *Let L_1 and L_2 be loops in $\mathcal{L}_2$ and K be a field whose characteristic does not divide the order of either of these loops. For $1 = 1, 2$, write $L_i = M_i \times A_i$ where M_i is 2-loop in $\mathcal{L}_2$ and A_i is an abelian group of odd order. Then $KL_1 \cong KL_2$ if and only if $K[M_1/M_1'] \cong K[M_2/M_2']$, $\Delta(M_1, M_1') \cong \Delta(M_2, M_2')$ and $KA_1 \cong KA_2$.*

Corollary 3.10. *Let L_1 and L_2 be loops in $\mathcal{L}_2$ and $\mathbf{Q}$ be the field of rationals. For $1 = 1, 2$, write $L_i = M_i \times A_i$ where M_i is 2-loop in $\mathcal{L}_2$ and A_i is an abelian group of odd order. Then $\mathbf{Q}L_1 \cong \mathbf{Q}L_2$ if and only if $M_1/M_1' \cong M_2/M_2'$, $\Delta(M_1, M_1') \cong \Delta(M_2, M_2')$ and $A_1 \cong A_2$.*

References

[1] **L. G. X. de Barros**: *Isomorphisms of rational loop algebras*, Comm. Algebra **21** (1993), 3977 – 3993.

[2] **L. G. X. de Barros**: *On semisimple alternative loop algebras*, Comm. Algebra **21** (1993), 3995 – 4011.

[3] **L. G. X. de Barros, C. Polcino Milies**: *Loop algebras of code loops*, Comm. Algebra **23** (1995), 4781 – 4790.

[4] **L. G. X. de Barros, C. Polcino Milies**: *Modular loop algebras of RA loops*, J. Algebra **23** (1995), 1027 – 1040.

[5] **O. Chein, E. G. Goodaire**: *Loops whose loop rings are alternative*, Comm. Algebra **175** (1986), 293 – 310.

[6] **O. Chein, E. G. Goodaire**: *Moufang loops with a unique non-identity commutator* (*associator, square*), J. Algebra **130** (1990), 369 – 384.

[7] **E. G. Goodaire, E. Jespers, C. Polcino Milies**, Alternative Loop Rings, North Holland Math. Series no. **184**, Elsevier, Amsterdam, 1996.

[8] **E. G. Goodaire, D. A. Robinson**: *A class of loops with right alternative loop rings*, Comm. Algebra **22** (1994), 5623 – 5634.

[9] **G. Leal, C. Polcino Milies**: *Isomorphic group* (*and loop*) *algebras*, J. Algebra **155** (1993), 195 – 210.

[10] **S. Perlis, G. L. Walker**: *Abelian group algebras of finite order*, Trans. Amer. Math. Soc. **68** (1950), 420 – 426.

Centro Universitário Senac
São Paulo, SP, CEP 04696-000
Brazil
e-mail: lbarros@sp.senac.br

Received June 27, 2005

Quasigroups and Related Systems **14** (2006), 11 – 26

Subassociative groupoids

Milton S. Braitt and Donald Silberger

Dedicated to the memory of Eva Ruth Silberger, 1962 – 2006

Abstract

When $\langle G;\diamond\rangle$ is a groupoid with binary operation $\diamond : G^2 \to G$, and when $k \in \mathbf{N} := \{1, 2, 3, \ldots\}$, then $F^\sigma(k)$ denotes the set of all formal products $\mathbf{u}$ on k independent variables. It is well known that $|F^\sigma(k)| = C(k)$, where $C(k)$ is the kth Catalan number.

Each word $\mathbf{u} \in F^\sigma(k)$ induces a function $\mathbf{u} : G^k \to G$ given by $\mathbf{u} : \vec{g} \mapsto \mathbf{u}(\diamond, \vec{g})$, where $\mathbf{u}(\diamond, \vec{g})$ is the interpretation in $\langle G;\diamond\rangle$ of $\mathbf{u}$ as a $\diamond$–product of the sequence $\vec{g} := \langle g_0, g_1, \ldots, g_{k-1}\rangle \in G^k$.

Write $\mathbf{u} =_\diamond \mathbf{v}$ for $\{\mathbf{u}, \mathbf{v}\} \subseteq F^\sigma(k)$ iff $\mathbf{u}(\diamond, \vec{g}) = \mathbf{v}(\diamond, \vec{g})$ whenever $\vec{g} \in G^k$. This $=_\diamond$ is an equivalence relation on the set $F^\sigma := \bigcup\{F^\sigma(k) : k \in \mathbf{N}\}$. The sequence $\mathbf{SaT}(\langle G;\diamond\rangle) := \langle |F^\sigma(k)/ =_\diamond| \rangle_{k=2}^\infty$ presents the subassociativity types of $\langle G;\diamond\rangle$.

We calculate $\mathbf{SaT}(G)$ for a few evocative groupoids $G := \langle G;\diamond\rangle$, and we initiate a study of the partitions $F^\sigma(k)/ =_\diamond$. Each equivalence class of the completely free groupoid F^σ is a singleton, and so F^σ realizes the theoretical minimum k–associativity for each $k \in \mathbf{N}$. We propose for each k a minimally k–associative class of finite groupoids.

Introduction

Given a set G and a *binary operation* $\diamond : G \times G \to G$ on G, it is customary to write $\diamond(x, y)$ in the form $x \diamond y$ when $\langle x, y\rangle \in G^2 := G \times G$. The pair $\langle G;\diamond\rangle$ is said to be a *groupoid*.

We say that a triple $\langle g_0, g_1, g_2\rangle \in G^3$ of elements in G *associates* under the binary operation $\diamond$ iff $(g_0 \diamond g_1) \diamond g_2 = g_0 \diamond (g_1 \diamond g_2)$. If every triple of

2000 Mathematics Subject Classification: 20N02, 08A99, 05A99
Keywords: groupoid, free groupoid, associative law

elements in G associates under $\diamond$, the binary operation $\diamond$ itself is said to be *associative*, and the groupoid $\langle G;\diamond\rangle$ is called a *semigroup*.

For $\langle G;\diamond\rangle$ a semigroup, each finite sequence $g_0, g_1, \ldots, g_{k-1}$ of elements in G determines under $\diamond$ a unique element in G as its product. We can write this product in the simplified form $g_0 \diamond g_1 \diamond \cdots \diamond g_{k-1}$ because parentheses are not needed to avoid ambiguity.

Of course, the great majority of groupoids are not semigroups. Each such nonsemigroup has at least one triple of elements which fails to associate. This failure of some triples to associate induces diversity among products of the longer strings as well.

Our paper's principal focus is upon this diversity of products.

For instance, if $\langle G;\diamond\rangle$ is not a semigroup then we expect that some quadruples $\langle g_0, g_1, g_2, g_3\rangle \in G^4$ may in a general sense also fail to associate. However, whereas there are at most two potentially distinct products for a triple of elements in a groupoid, there are five potentially distinct products of a quadruple of such elements, fourteen potentially distinct 5–tuple products, and in general there are $C(k)$ potentially distinct k–products, where $C(k)$ is the kth Catalan number.

That is, when the binary operation $\diamond$ of a groupoid lacks 3–associativity, then $\diamond$ may lack k–associativity for sundry integers $k \geqslant 4$ as well.

In §1 we introduce Reverse Polish Notation, which provides a convenient tool for specifying the potentially different k–products under $\diamond$ of a length k sequence of elements in G. This leads to our presentation in §2 of the notion of a formal k–product, and of a completely free groupoid in which every formal k–product of a length k sequence in G produces a *de facto* distinct element in the groupoid, and enables our development in §3 of a measure of the subassociativity of an arbitrary groupoid; this measure is given as an infinite sequence of positive integers which we call the subassociativity type of the groupoid. In §3 we calculate the subassociativity type of each of several important nonassociative groupoids, including that of the groupoid of integers under subtraction. Related to the subassociativity type of a groupoid is its size sequence, which appears interestingly complicated even when the subassociativity type of the groupoid is regular and simple in form.

§4 concentrates upon those groupoids in which k–associativity is minimal for every integer $k \geqslant 3$.

Our paper presents a variety of natural problems.

1. Reverse Polish Notation

For a nonassociative binary operation $\diamond$, if $\vec{g} := \langle g_0, g_1, \ldots, g_{k-1}\rangle$ is a finite sequence of elements in G then it may happen that $\mathbf{w}(\diamond, \vec{g}) \neq \mathbf{v}(\diamond, \vec{g})$, where $\mathbf{w}$ and $\mathbf{v}$ are some two "appropriate parenthesizations" of the augmented sequence $\diamond(\vec{g}) := \langle g_0, \diamond, g_1, \diamond, \cdots, \diamond, g_{k-1}\rangle$.

We call a parenthesization of $\diamond(\vec{g})$ *appropriate* if it enables the $k-1$ occurrences of the symbol $\diamond$ unambiguously to serve as a binary operation in $\diamond(\vec{g})$. For instance, the two parenthesizations in (1), below, are appropriate; and, if $\diamond$ is associative, then we can believe that

$$((g_0 \diamond (g_1 \diamond g_2)) \diamond (g_3 \diamond g_4)) = ((g_0 \diamond g_1) \diamond ((g_2 \diamond g_3) \diamond g_4)). \tag{1}$$

We have enclosed each of the two expressions, balanced by the $=$ sign in (1), with an external, conventionally unnecessary, parenthesis pair, whose purpose is to assure that each $\diamond$–multiplication is consistent in its form; namely, $(a \diamond b)$, instead of $a \diamond b$ as sometimes abbreviated. Our reason for this ostensible redundancy of parentheses should become clear after our discussion, in the next few paragraphs, of Reverse Polish Notation (RPN).

RPN is sometimes more convenient than parenthesized expressions of the sort in (1). For people who are uneasy with RPN we provide a gradual approach to this, parenthesis-free, notation. In two steps we will convert the usual-form equality (1) into its equivalent RPN version, (3).

First, we remove left parentheses from (1). A routine proof shows that there is exactly one way to restore left parentheses to the resulting left-parenthesis deprived expression, (2), so as to regain an appropriately parenthesized augmented sequence. Here then is (2):

$$g_0 \diamond g_1 \diamond g_2)) \diamond g_3 \diamond g_4)) = g_0 \diamond g_1) \diamond g_2 \diamond g_3) \diamond g_4)) \tag{2}$$

Each of the $=$ expressions in (1) and (2) has 5 terms, g_i, which are elements in G. It is no accident that each of those expressions has also exactly 4 occurrences of $\diamond$ and exactly 4 right parentheses. In order to create from the expressions in (2) their equivalent RPN expressions we merely eliminate the 4 occurrences in (2) of $\diamond$, and then in the $\diamond$–free resulting expression we replace each right parenthesis with a new occurrence of $\diamond$. Thus, finally, we obtain the RPN equation which is equivalent to (1):

$$g_0 g_1 g_2 \diamond \diamond g_3 g_4 \diamond \diamond = g_0 g_1 \diamond g_2 g_3 \diamond g_4 \diamond \diamond. \tag{3}$$

Comparing (1) and (3), we see that (3) is shorter than (1). RPN is an efficient way of representing lengthy $\diamond$–products. One can safely remove all parentheses from (1) and maintain a *bona fide* equality if and only if $\diamond$ is associative. But with the RPN expression, (3), there are no parentheses to remove, and when $\diamond$ is associative then every RPN product constructed from the sequence $\langle g_0, g_1, g_2, g_3, g_4\rangle \in G^5$ is equal to that given by

$$g_0 g_1 g_2 g_3 g_4 \diamond^4 .$$

This too is shorter than the usual product expression

$$g_0 \diamond g_1 \diamond g_2 \diamond g_3 \diamond g_4.$$

RPN confers a more important advantage: It facilitates our classification of the "subassociativity" of groupoids.

2. Formal Products

We will define a groupoid, F^σ, inspired by the idea of the "completely free" groupoid $F := F(x, \bullet)$ generated by a two-letter alphabet, $\{x, \bullet\}$.

$F \subset \{x, \bullet\}^*$, where $\{x, \bullet\}^*$ is the semigroup under concatenation of all finite words with letters in $\{x, \bullet\}$.

We write $\mathbf{a} = \mathbf{b}$ to say that the word $\mathbf{a}$ is spelled the same as the word $\mathbf{b}$, for $\{\mathbf{a}, \mathbf{b}\} \subseteq F$.

$\#(\mathbf{u}, z)$ denotes the number of occurrences of a letter z in the word $\mathbf{u}$. A nonempty word $\mathbf{w} \in \{\bullet, x\}^*$ is an element in F if and only if

(i) $\#(\mathbf{w}, x) - \#(\mathbf{w}, \bullet) = 1$.

(ii) If $\mathbf{p}$ is a nonempty prefix of $\mathbf{w}$ then $\#(\mathbf{p}, x) > \#(\mathbf{p}, \bullet)$.

It is easy to see that $x \in F$, and that $\{\mathbf{u}, \mathbf{v}\} \subseteq F \Rightarrow \mathbf{uv}\bullet \in F$. Thus $\bullet$ serves in F as an operator symbol, providing a binary operation $\bullet : \langle \mathbf{u}, \mathbf{v}\rangle \mapsto \mathbf{uv}\bullet$ for F in RPN format.

The relevant property of the groupoid $\langle F, \bullet\rangle$ is that if $\langle \mathbf{p}, \mathbf{s}\rangle \neq \langle \mathbf{p}', \mathbf{s}'\rangle$ with $\{\mathbf{p}, \mathbf{p}', \mathbf{s}, \mathbf{s}'\} \subset F$ then $\mathbf{ps} \neq \mathbf{p}'\mathbf{s}'$. An easy related fact is that the binary operation $\bullet$ is antiassociative; *i.e.*, that no triples in F associate:

Theorem 2.1. *Let* $\langle \mathbf{a}, \mathbf{b}, \mathbf{c}\rangle \in F^3$. *Then* $\mathbf{abc}\bullet\bullet \neq \mathbf{ab}\bullet\mathbf{c}\bullet$.

Proof. Clearly the "free" semigroup $\{x, \bullet\}^*$ has the cancellation property. Thus, if $\mathbf{abc}\bullet\bullet = \mathbf{ab}\bullet\mathbf{c}\bullet$ then $\mathbf{c}\bullet\bullet = \bullet\mathbf{c}\bullet$. But $\mathbf{c}\bullet\bullet \neq \bullet\mathbf{c}\bullet$, since x is a prefix of $\mathbf{c}$ and since $x \neq \bullet$. □

Notice that, if $\mathbf{w} \in F$, then either $\mathbf{w} = x$ or there exists exactly one pair $\langle \mathbf{p}, \mathbf{s} \rangle \in F \times F$ such that $\mathbf{w} = \mathbf{ps}\bullet$.

Henceforth $\vec{x} := \langle x_0, x_1, x_2, \ldots \rangle$ is a sequence of distinct variables, and $\bullet$ is an operator symbol. Let k be a positive integer. We now modify F:

Definition 2.2. By a *formal k–product* we mean any word $\mathbf{w}$ of length $2k-1$ in the alphabet $\{x_0, x_1, \ldots, x_{k-1}, \bullet\}$, satisfying three conditions:

(i) $x_0 x_1 \ldots x_{k-1}$ is a subword of $\mathbf{w}$.

(ii) $\mathbf{w}$ has exactly $k-1$ occurrences of the operator symbol $\bullet$.

(iii) If $\mathbf{p}$ is a nonempty prefix of $\mathbf{w}$ then $\mathbf{p}$ has fewer occurrences of the operator symbol $\bullet$ than it has of variable symbols x_i.

As usual $\mathbf{N} := \{1, 2, 3, \ldots\}$. For $k \in \mathbf{N}$ the expression $F^\sigma(k)$ denotes the set of all formal k–products. Finally, we define the infinite set F^σ by

$$F^\sigma := \bigcup \{F^\sigma(k) : k \in \mathbf{N}\}.$$

It is well-known, *viz* [1, 2, 3, 4], that for each $k \in \mathbf{N}$ the number $|F^\sigma(k)|$ is the kth term of the Catalan sequence, which is to say that

$$|F^\sigma(k)| = C(k) := \frac{1}{2k-1}\binom{2k-1}{k}.$$

Henceforth $\omega := \mathbf{N} \cup \{0\}$, and $k := \{0, 1, \ldots, k-1\}$ when $k \in \mathbf{N}$.

For $\langle k, j \rangle \in \mathbf{N} \times \omega$ and $\mathbf{w} \in F^\sigma(k)$, the expression $\mathbf{w}_j$ denotes the word obtained by replacing the letter x_i in $\mathbf{w}$ with the letter x_{j+i} for each $i \in k$. We write $F^\sigma_j(k) := \{\mathbf{u}_j : \mathbf{u} \in F^\sigma(k)\}$.

Illustrative Example 1: When $\mathbf{w} := x_0 x_1 \bullet x_2 x_3 \bullet x_4 \bullet\bullet$ then $\mathbf{w} \in F^\sigma(5)$, and $\mathbf{w}_{13} = x_{13}x_{14} \bullet x_{15}x_{16} \bullet x_{17} \bullet\bullet \in F^\sigma_{13}(5)$.

Observe that $\langle F^\sigma; \odot \rangle$ is a groupoid where the binary operation $\odot$ is defined thus: When $\langle \mathbf{u}, \mathbf{v} \rangle \in F^\sigma(k) \times F^\sigma(j)$, then $\mathbf{uv}\odot := \mathbf{uv}_k\bullet$. It is easy to see that then $\mathbf{uv}\odot \in F^\sigma(k+j)$.

Indeed, $F^\sigma(1) = \{x_0\}$, while for $2 \leqslant k \in \mathbf{N}$ one could show that

$$F^\sigma(k) = \bigcup\{F^\sigma(i)F^\sigma(k-i)\odot : \ 1 \leqslant i \leqslant k-1\},$$

where $F^\sigma(i)F^\sigma(k-i)\odot := \{\mathbf{uv}\odot : \ \mathbf{u} \in F^\sigma(i) \wedge \mathbf{v} \in F^\sigma(k-i)\}$.

Illustrative Example 2: When $\mathbf{u} := x_0x_1 \bullet x_2x_3 \bullet\bullet \in F^\sigma(4)$, and when $\mathbf{v} := x_0x_1x_2x_3 \bullet x_4 \bullet\bullet\bullet \in F^\sigma(5)$, we have that

$$\begin{aligned}\mathbf{uv}\odot &= x_0x_1 \bullet x_2x_3 \bullet\bullet \ \ x_0x_1x_2x_3 \bullet x_4 \bullet\bullet\bullet \odot \\ &= x_0x_1 \bullet x_2x_3 \bullet\bullet x_{4+0}x_{4+1}x_{4+2}x_{4+3} \bullet x_{4+4} \bullet\bullet\bullet\bullet \\ &= x_0x_1 \bullet x_2x_3 \bullet\bullet x_4x_5x_6x_7 \bullet x_8 \bullet\bullet\bullet\bullet \in F^\sigma(4+5)\,.\end{aligned}$$

Theorem 2.3. *If* $\langle \mathbf{a}, \mathbf{b}, \mathbf{c}\rangle \in (F^\sigma)^3$ *then* $\mathbf{ab}\odot\mathbf{c}\odot \neq \mathbf{abc}\odot\odot$.

Proof. $\langle \mathbf{a}, \mathbf{b}, \mathbf{c}\rangle \in F^\sigma(i) \times F^\sigma(j) \times F^\sigma(t)$ for some $\langle i, j, t\rangle \in \mathbf{N}^3$. Thus $\mathbf{ab}\odot\mathbf{c}\odot = \mathbf{ab}_i \bullet \mathbf{c}\odot = \mathbf{ab}_i \bullet \mathbf{c}_{i+j}\bullet$ and $\mathbf{abc}\odot\odot = \mathbf{abc}_j \bullet \odot = \mathbf{ab}_i\mathbf{c}_{i+j} \bullet\bullet$. So if $\mathbf{ab}\odot\mathbf{c}\odot = \mathbf{abc}\odot\odot$ then $\bullet\,\mathbf{c}_{i+j}\bullet = \mathbf{c}_{i+j}\bullet\bullet$, an impossibility. □

In our view, the groupoids $\langle F; \bullet\rangle$ and $\langle F^\sigma; \odot\rangle$ lie at an opposite extreme from the class of semigroups. For, no triple of elements in either of these two groupoids associates. However, every triple in a semigroup associates.

For $\langle G; \diamond\rangle$ a semigroup, each sequence $\vec{g} := \langle g_0, g_1, \ldots, g_{k-1}\rangle \in G^k$ unambiguously determines the element in G obtained by the conventionally presented but unparenthesized product $g_0 \diamond g_1 \diamond \cdots \diamond g_{k-1}$.

If we endowed F^σ with the relation, $x_0x_0 \odot x_0 \odot \approx x_0x_0x_0 \odot \odot$, then $|F^\sigma(k)/\approx| = 1$ for each $k \in \mathbf{N}$, where $F^\sigma(k)/\approx$ is the family of $\approx$–equivalence classes.

Plainly every groupoid falls between the extremes represented by $\langle F^\sigma; \odot\rangle$ on one end, and by the class of semigroups on the other. We believe that every finite nonassociative groupoid lies strictly between these extremes.

We next propose a scheme for using F^σ in order to pin down this idea.

3. The Subassociativity Type of a Groupoid

Let $\langle G; \diamond\rangle$ be an arbitrary groupoid, let $\mathbf{w} \in F^\sigma(k)$ for a given $k \in \mathbf{N}$, and let $\vec{g} := \langle g_0, g_1, g_2, \ldots\rangle \in G^\infty$. Then $\mathbf{w}(\diamond, \vec{g})$ denotes the element in G obtained by replacing in $\mathbf{w}$ the operator $\bullet$ with the operation $\diamond$, and the

symbol x_i with the element g_i for each $i \in k$, and then executing the $k-1$ binary operations $\diamond$ as indicated in the modified version of $\mathbf{w}$.

Illustrative Example 3: Consider the groupoid $\langle \mathbf{Z}; - \rangle$ of integers under subtraction, and the formal 5–product $\mathbf{w} := x_0x_1 \bullet x_2x_3x_4 \bullet\bullet\bullet \in F^\sigma(5)$. Let $\vec{g} := \langle 2, 7, 0, 1, -5, g_5, g_6, \ldots \rangle$. Then $\mathbf{w}(-, \vec{g}) = 2\,7\,-\,0\,1\,(-5)\,-\,--$, where we append parentheses to eliminate ambiguity. In conventional form $\mathbf{w}(-, \vec{g}) = (2-7)-(0-(1-(-5)))$, and hence $\mathbf{w}(-, \vec{g}) = 1$.

Definition 3.1. Let $k \geqslant 3$, and let $\{\mathbf{u}, \mathbf{v}\} \subseteq F^\sigma(k)$. Let $\langle G; \diamond \rangle$ be a groupoid. We say that $\mathbf{u}$ is $\diamond$–*equivalent* to $\mathbf{v}$, in which event we write $\mathbf{u} \approx_\diamond \mathbf{v}$, iff $\mathbf{u}(\diamond, \vec{g}) = \mathbf{v}(\diamond, \vec{g})$ for all $\vec{g} \in G^\infty$. The expressions $F^\sigma(k)/\diamond$ and $F^\sigma(k)/\approx_\diamond$ denote the family of $\approx_\diamond$–equivalence classes $[\mathbf{w}]_\diamond$ of $F^\sigma(k)$.

Definition 3.2. We call a groupoid $\langle G; \diamond \rangle$ *completely free* iff $[\mathbf{w}]_\diamond = \{\mathbf{w}\}$ for every $\mathbf{w} \in F^\sigma$.

Illustrative Example 4: Returning to the groupoid $\langle \mathbf{Z}; - \rangle$ of Example 3, we easily see for $k \in \{1, 2, 3\}$ that $F^\sigma(k)/\approx_- = \{\{\mathbf{w}\} : \mathbf{w} \in F^\sigma(k)\}$. But for $k = 4$ the situation complicates slightly. As we will proceed to show, $\mathbf{u} \approx_- \mathbf{v}$ where $\mathbf{u} := x_0x_1x_2x_3 \bullet\bullet\bullet$ and where $\mathbf{v} := x_0x_1x_2 \bullet\bullet x_3 \bullet$:

Switching back and forth between RPN and ordinary terminology as convenience dictates, we note for an arbitrary $\vec{g} \in \mathbf{Z}^\infty$ that $\mathbf{u}(-, \vec{g}) = g_0g_1g_2g_3--- = g_0-(g_1-(g_2-g_3)) = g_0-g_1+g_2-g_3 = (g_0-(g_1-g_2))-g_3 = g_0g_1g_2--g_3- = \mathbf{v}(-, \vec{g})$. Similar calculations establish that $F^\sigma(4)/\approx_- = \{\{\mathbf{u}, \mathbf{v}\}, \{\mathbf{a}\}, \{\mathbf{b}\}, \{\mathbf{c}\}\}$, where $\mathbf{a} := x_0x_1x_2 \bullet x_3 \bullet\bullet$ and $\mathbf{b} := x_0x_1 \bullet x_2x_3 \bullet\bullet$ and $\mathbf{c} := x_0x_1 \bullet x_2 \bullet x_3 \bullet$. Thus $|F^\sigma(4)/\approx_-| = 4 < 5 = |F^\sigma(4)|$. So the groupoid $\langle \mathbf{Z}; - \rangle$ is neither a semigroup, nor is it completely free.

Theorem 3.3. *Let $\langle G; \diamond \rangle$ be a groupoid, and let $\{\mathbf{s}, \mathbf{t}, \mathbf{s}', \mathbf{t}'\} \subset F^\sigma$. Let $\mathbf{s} \approx_\diamond \mathbf{s}'$, and let $\mathbf{t} \approx_\diamond \mathbf{t}'$. Then $\mathbf{st}\odot \approx_\diamond \mathbf{s}'\mathbf{t}'\odot$.*

Proof. There exist $k \in \mathbf{N}$ such that $\{\mathbf{s}, \mathbf{s}'\} \subseteq F^\sigma(k)$, and $j \in \mathbf{N}$ such that $\{\mathbf{t}, \mathbf{t}'\} \subseteq F^\sigma(j)$. Pick $\vec{g} := \langle g_0, g_1, \ldots, g_{k-1}, g_k, \ldots, g_{k+j-1}, \ldots \rangle \in G^\infty$. Let $\vec{b} := \langle g_k, g_{k+1}, \ldots, g_{k+j-1}, \ldots \rangle \in G^\infty$.

By hypothesis $\mathbf{s}(\diamond, \vec{g}) = \mathbf{s}'(\diamond, \vec{g})$ and $\mathbf{t}(\diamond, \vec{b}) = \mathbf{t}'(\diamond, \vec{b})$. Therefore

$$(\mathbf{st}\odot)(\diamond, \vec{g}) = (\mathbf{st_k}\bullet)(\diamond, \vec{g}) = [\mathbf{s}(\diamond, \vec{g})][\mathbf{t}_k(\diamond, \vec{g})]\diamond = [\mathbf{s}(\diamond, \vec{g})][\mathbf{t}(\diamond, \vec{b})]\diamond =$$

$$[\mathbf{s}'(\diamond, \vec{g})][\mathbf{t}'(\diamond, \vec{b})]\diamond = [\mathbf{s}'(\diamond, \vec{g})][\mathbf{t}'_k(\diamond, \vec{g})]\diamond = (\mathbf{s}'\mathbf{t}'_k\bullet)(\diamond, \vec{g}) = (\mathbf{s}'\mathbf{t}'\odot)(\diamond, \vec{g}),$$

with parentheses and brackets appended only to aid the reader. □

Recall that $C(k)$ denotes the kth Catalan number. The following is an easy consequence of Theorem 3.3 and Illustrative Example 4.

Corollary 3.4. *If* $k \in \{1,2,3\}$ *then* $|F^\sigma(k)/\approx_- \,| = C(k)$. *However,* $|F^\sigma(j)/\approx_- \,| < C(j)$ *for every integer* $j \geqslant 4$.

Definition 3.5. For $\langle G;\diamond\rangle$ a groupoid, we define the *subassociativity type* of this groupoid to be the infinite sequence in $\mathbf{N}$, written

$$\mathbf{SaT}(\langle G;\diamond\rangle) := \langle\, |F^\sigma(k)/\approx_\diamond| \,\rangle_{k=2}^{\infty} .$$

For $\langle S;\cdot\,\rangle$ a semigroup, obviously $\mathbf{SaT}(\langle S;\cdot\,\rangle) = \langle 1,1,1,\ldots\rangle$.

As we remarked, $\mathbf{SaT}(\langle F^\sigma;\odot\rangle) = \langle C(n)\rangle_{n=2}^{\infty}$.

Theorem 3.6. $\mathbf{SaT}(\langle \mathbf{Z};-\rangle) = \langle 2^{k-2}\rangle_{k=2}^{\infty}$.

Proof. For each integer $k \geqslant 2$, for each $\mathbf{w} \in F^\sigma(k)$, and for each $\vec{g} \in \mathbf{Z}^k$, we observe that $\mathbf{w}(-,\vec{g}) = g_0 - g_1 \pm_1 g_2 \pm_2 g_3 \pm_3 \cdots \pm_{k-3} g_{k-2} \pm_{k-2} g_{k-1}$ for some "sign" sequence $\langle \pm_1, \pm_2, \ldots, \pm_{k-2}\rangle \in \{-,+\}^{k-2}$, where we present the expression to the right of the symbol $=$ in ordinary terminology. Indeed, since there are only 2^{k-2} distinct sign sequences of length $k-2$, we see that $|F^\sigma(k)/\approx_- \,| \leq 2^{k-2}$. So it suffices to show that $|F^\sigma(k)/\approx_- \,| \not< 2^{k-2}$.

Claim: For every sign sequence $\langle \pm_1, \pm_2, \ldots, \pm_{k-2}\rangle \in \{-,+\}^{k-2}$, there exists $\mathbf{r} \in F^\sigma(k)$ such that $\mathbf{r}(-,\vec{g}) = g_0 - g_1 \pm_1 g_2 \pm_2 \cdots g_{k-2} \pm_{k-2} g_{k-1}$, thus "realizing" the sign sequence $\langle \pm_i\rangle_{i=1}^{k-2}$.

We argue by induction on $k \geqslant 2$. The claim is trivial for $k = 2$.

For $2 \leqslant n \in \mathbf{N}$, suppose the claim holds when $k = n$. Pick a sign sequence $\langle \pm_i\rangle_{i=1}^{n-2} \in \{-1,1\}^{n-2}$ and a sequence $\vec{h} := \langle h_0, h_1, \ldots, h_n\rangle \in \mathbf{Z}^{n+1}$. We are required only to supply $\{\mathbf{w}_-, \mathbf{w}_+\} \subset F^\sigma(n+1)$ such that

$$\mathbf{w}_-(-,\vec{h}) = h_0 - h_1 \pm_1 h_2 \pm_2 \cdots \pm_{n-3} h_{n-2} \pm_{n-2} h_{n-1} - h_n$$

and such that

$$\mathbf{w}_+(-,\vec{h}) = h_0 - h_1 \pm_1 h_2 \pm_2 \cdots \pm_{n-3} h_{n-2} \pm_{n-2} h_{n-1} + h_n.$$

For each positive integer $i \leqslant n-2$ let $\mp_i := -\pm_i$. And now define $\vec{p} := \langle h_0, h_1, h_2, \ldots, h_{n-2}, h_{n-1}, \ldots\rangle$ and $\vec{s} := \langle h_1, \pm_1 h_2, h_3, \ldots, h_{n-1}, h_n, \ldots\rangle$. Both $\vec{p}$ and $\vec{s}$ are sequences in $\mathbf{Z}^\infty$.

By the inductive hypothesis there exists $\mathbf{u} \in F^\sigma(n)$ such that

$$\mathbf{u}(-,\vec{p}) = h_0 - h_1 \pm_1 h_2 \pm_2 \cdots h_{n-2} \pm_{n-2} h_{n-1}.$$

Let $\mathbf{w}_- := \mathbf{u}x_0\odot$. Then, with ordinary terminology when convenient,

$$\mathbf{w}_-(-,\vec{h}) = (\mathbf{u}x_n\bullet)(-,\vec{h}) = (\mathbf{u}(-,\vec{p}))-h_n = h_0-h_1\pm_1\cdots\pm_{n-2}h_{n-1}-h_n\,.$$

Again by the inductive hypothesis there exists $\mathbf{v}\in F^\sigma(n)$ such that

$$\mathbf{v}(-,\vec{s}) = h_1 - (\pm_1 h_2)\mp_2 h_3\mp_3\cdots\mp_{n-2}h_{n-1} - h_n.$$

Let $\mathbf{w}_+ := x_0\mathbf{v}\odot$. Then

$$\mathbf{w}_+(-,\vec{h}) = (x_0\mathbf{v}_1\bullet)(-,\vec{h}) = h_0 - (\mathbf{v}(-,\vec{s})) =$$

$$h_0 - (h_1 - (\pm_1 h_2)\mp_2 h_3\mp_3\cdots\mp_{n-2}h_{n-1} - h_n) =$$

$$h_0 - h_1\pm_1 h_2\pm_2 h_3\pm_3\cdots\pm_{n-2}h_{n-1} + h_n.$$

The theorem follows. □

Corollary 3.4 applies to each groupoid $\langle G;\diamond\rangle$. If $|F^\sigma(k)/\approx_\diamond| < C(k)$ then $|F^\sigma(j)/\approx_\diamond| < C(j)$ for all $j > k$.

The subtraction of integers is an issue for the very young. Surely one ought to be able to settle every relevant question about the groupoid $\langle\mathbf{Z};-\rangle$.

Theorem 3.6 suggests further scrutiny. Since the average size

$$\frac{C(k)}{2^{k-2}}$$

of the equivalence classes $[\mathbf{w}]_-\in F^\sigma(k)/\approx_-$ increases without bound as k increases, it is reasonable to wonder how the sizes of those equivalence classes are distributed

Definition 3.7. For each $k\geqslant 2$ we say that the sequence $\langle\langle\nu_k(i),i\rangle\rangle_{i=1}^\infty$ in $\omega\times\mathbf{N}$ is the *size sequence* for k *of* $\langle\mathbf{Z};-\rangle$ when $F^\sigma(k)/\approx_-$ contains exactly $\nu_k(i)$ member sets of size i, for each $i\in\mathbf{N}$.

Of course in any size sequence, $\nu_k(i) > 0$ for only finitely many i. For a size sequence we list only those terms with positive first coordinates.

Here are the size sequences and other relevant numerical data about $F^\sigma(k)/\approx_-$ for the cases $k\in\{4,5,6\}$:

$|F^\sigma(4)| = 5$ and $|F^\sigma(4)/\approx_-| = 4$. Size sequence: $\langle 3,1\rangle$, $\langle 1,2\rangle$.

$|F^\sigma(5)| = 14$ and $|F^\sigma(5)/\approx_-| = 8$. Size sequence: $\langle 4,1\rangle$, $\langle 3,2\rangle$, $\langle 1,4\rangle$.

$|F^\sigma(6)| = 42$ and $|F^\sigma(6)/\approx_-| = 16$. Size sequence: $\langle 5,1\rangle$, $\langle 3,2\rangle$, $\langle 4,3\rangle$, $\langle 2,4\rangle$, $\langle 1,5\rangle$, $\langle 1,6\rangle$.

Problem 3.8. *Specify the size sequences of* $\langle \mathbf{Z}; - \rangle$ *for each* $k \geqslant 2$.

Observe that a groupoid $\langle G; \diamond \rangle$ is completely free if and only if $|F^\sigma(k)| = |F^\sigma(k)/\approx_\diamond|$ for all $k \in \mathbf{N}$.

Suggestive Example 5: Let the binary operation $\triangleleft$ on the set $2 := \{0, 1\}$ be given by $0t\triangleleft := 1$ and $1t\triangleleft := 0$ for each $t \in 2$.

It is easily checked that the groupoid, $\langle 2; \triangleleft \rangle$, is antiassociative. Hence, $F^\sigma(3)/\approx_\triangleleft = \{\{x_0x_1x_2 \bullet\bullet\}, \{x_0x_1 \bullet x_2 \bullet\}\}$. $F^\sigma(4)/\approx_\triangleleft := \{A, B\}$ where $|A| = 3$ and $|B| = 2$. In fact

$$A = \{x_0x_1 \bullet x_2 \bullet x_3 \bullet,\ x_0x_1x_2 \bullet x_3 \bullet\bullet,\ x_0x_1x_2x_3 \bullet\bullet\bullet\},$$
$$B = \{x_0x_1x_2 \bullet\bullet x_3 \bullet,\ x_0x_1 \bullet x_2x_3 \bullet\bullet\}.$$

$F^\sigma(5)/\approx_\triangleleft = \{C, D\}$, where C contains 8 of the elements in $F^\sigma(5)$ while D contains the other 6 formal 5–products.

Theorem 3.9. $|F^\sigma(k)/\approx_\triangleleft| = 2$ *for all* $k \geqslant 3$.

Proof. Choose $k \geqslant 3$. Our "test sequence" is $\vec{0} := \langle 0, 0, \ldots \rangle$. For $\mathbf{w} \in F^\sigma(k)$ we write $\mathbf{w} \in A_k$ to mean that $\mathbf{w}(\triangleleft, \vec{0}) = 0$, and we define $B_k := F^\sigma(k) \setminus A_k$.

For every positive integer pair $\langle i, j \rangle$ such that $i + j = k$, we have that $A_iF^\sigma(j)\odot \subset B_k$. Indeed

$$\bigcup_{i=1}^{k-1} A_iF^\sigma(k-i)\odot = B_k$$

and similarly

$$\bigcup_{i=1}^{k-1} B_iF^\sigma(k-i)\odot = A_k.$$

So $\{A_k, B_k\}$ is a partition of $F^\sigma(k)$, since

$$F^\sigma(k) = \bigcup_{i=1}^{k-1} F^\sigma(i)F^\sigma(k-i)\odot = \bigcup_{i=1}^{k-1} (A_i \dot\cup B_i)F^\sigma(k-i)\odot = B_k \dot\cup A_k.$$

It remains only to show that we were wise in our choice of $\vec{0}$ as a test sequence in 2^∞. For an arbitrary pair $\mathbf{u}$ and $\mathbf{v}$ of formal k–products we must prove that $\mathbf{u}(\triangleleft, \vec{0}) = \mathbf{v}(\triangleleft, \vec{0}) \Rightarrow \mathbf{u} \approx_\triangleleft \mathbf{v}$. So, pick any $\vec{g} \in 2^\infty$.

Recall that $\mathbf{u} = \mathbf{p}_0\mathbf{s}_0\odot$ for a unique pair $\langle \mathbf{p}_0, \mathbf{s}_0 \rangle \in F^\sigma \times F^\sigma$. Likewise there is a unique $\langle \mathbf{p}_1, \mathbf{s}_1 \rangle \in F^\sigma \times F^\sigma$ with $\mathbf{p}_0 = \mathbf{p}_1\mathbf{s}_1\odot$. Proceeding, we

obtain a unique descending sequence of formal product prefixes of $\mathbf{u}$. There exists $\lambda := \lambda(\mathbf{u}) \in \mathbf{N}$ for which $\mathbf{p}_\lambda$ is the final and shortest term $\neq x_0$ of the sequence. Notice that $\mathbf{p}_\lambda(\triangleleft, \vec{g}) = \overline{g_0}$ where $\overline{0} := 1$ and $\overline{1} = 0$. Furthermore,

$$\mathbf{u} = \mathbf{p}_\lambda \mathbf{s}_\lambda \odot \mathbf{s}_{\lambda-1} \odot \cdots \odot \mathbf{s}_1 \odot \mathbf{s}_0 \odot .$$

Thus we see that $\mathbf{u}(\triangleleft, \vec{g}) = g_0$ if the integer $\lambda(\mathbf{u})$ is even, but that $\mathbf{u}(\triangleleft, \vec{g}) = \overline{g_0}$ if $\lambda(\mathbf{u})$ is odd. A parallel analysis holds for $\mathbf{v}$. Thus $\mathbf{u}(\triangleleft, \vec{g}) = \mathbf{v}(\triangleleft, \vec{g})$ if and only if $\mathbf{u}(\triangleleft, \vec{0}) = \mathbf{v}(\triangleleft, \vec{0})$.

So now, for $\{\mathbf{u}, \mathbf{v}\} \subseteq F^\sigma(k)$, we see that

$$\mathbf{u} \approx_\triangleleft \mathbf{v} \;\Leftrightarrow\; \mathbf{u}(\triangleleft, \vec{0}) = \mathbf{v}(\triangleleft, \vec{0}) \;\Leftrightarrow\; \lambda(\mathbf{u}) + \lambda(\mathbf{v}) \quad \text{is even} \;\Leftrightarrow$$

$$(\{\mathbf{u}, \mathbf{v}\} \subseteq A_k \vee \{\mathbf{u}, \mathbf{v}\} \subseteq B_k) .$$

The theorem follows. □

Theorem 3.9 generalizes to an infinite class of antiassociative groupoids $\langle n; \hat{\delta} \rangle$ for $2 \leqslant n \in \mathbf{N}$. Indeed, the groupoid in Theorem 3.9 is the smallest example of the sort we will call "vertically deranged".

The expression $\mathrm{Sym}(G)$ denotes the collection of all permutations on the set G. And $\mathrm{Drn}(G)$ denotes the set of all *derangements* of G, which are those $f \in \mathrm{Sym}(G)$ such that $x \neq xf$ for every $x \in G$, where xf denotes the image — often written $f(x)$ — of x under the function f.

We call a groupoid $\langle G; \hat{\delta} \rangle$ *vertically deranged* if there is a derangement $\delta \in \mathrm{Drn}(n)$ such that $xy\hat{\delta} := x\delta$ for every $\langle x, y \rangle \in n \times n$, and we say that δ induces $\hat{\delta}$. Remarks analogous to those below apply to the horizontally deranged groupoid $\langle G; \check{\delta} \rangle$ where $xy\check{\delta} := y\delta$.

Theorem 3.10. *Every vertically deranged groupoid is antiassociative.*

Proof. Let $\langle G; \hat{\delta} \rangle$ be vertically deranged *via* some $\delta \in \mathrm{Drn}(G)$. Let $\vec{g} \in G^3$. Then $g_0 g_1 g_2 \hat{\delta}\hat{\delta} = g_0 g_1 \delta\hat{\delta} = g_0\delta \neq g_0\delta\delta = g_0 g_1 \hat{\delta}\delta = g_0 g_1 \hat{\delta} g_2 \hat{\delta}$. □

In the interest of maximizing the size of the families $F^\sigma(k)/\approx_\diamond$ induced by the groupoids $\langle n; \diamond \rangle$ for a fixed $n \in \mathbf{N}$, it seems prudent first to consider those $\langle n; \diamond \rangle$ which are antiassociative. The vertically deranged $\langle n; \diamond \rangle$ constitute a convenient class of finite antiassociative groupoids.

If $2 \leqslant n \in \mathbf{N}$ and if δ is a cyclic permutation of n, then for $i \in \omega$ it is evident that $\delta^i \in \mathrm{Drn}(n)$ if and only if i is not a multiple of n.

Theorem 3.11. *Let $2 \leqslant n \in \mathbf{N}$. Let $\delta \in Sym(n)$ be cyclic. Then*

$$|F^\sigma(k)/\approx_{\hat{\delta}}| = \min\{k-1, n\}.$$

Proof. By Theorem 3.9 we can take it that $n \geqslant 3$.

Let $k \in \{2, \ldots, n\}$. Pick $\mathbf{w} \in F^\sigma(k)$ and $\vec{g} := \langle g_0, g_1, g_2, \ldots\rangle \in n^\infty$.

Claim One: $\mathbf{w}(\hat{\delta}, \vec{g}) = g_0\delta^i$ for some positive integer $i < k$.

First, for $k = 2$ observe that $(x_0x_0\odot)(\hat{\delta}, \vec{g}) = g_0g_1\hat{\delta} := g_0\delta =: g_0\delta^1$.

Choose $k \geqslant 3$. Suppose that whenever $2 \leqslant j < k$,

$$\mathbf{u} \in F^\sigma(j) \Rightarrow \mathbf{u}(\hat{\delta}, \vec{g}) = g_0\delta^t$$

for some $t \in \{1, 2, \ldots, j-1\}$. Factor $\mathbf{w}$ in $\langle F^\sigma, \odot\rangle$: $\mathbf{w} = \mathbf{ps}\odot$. Since $\mathbf{p} \in F^\sigma(j)$ for some $j < k$, by hypothesis there exists t with $1 \leqslant t \leqslant j-1$ such that $\mathbf{p}(\hat{\delta}, \vec{g}) = g_0\delta^t$. From earlier calculations, $(\mathbf{ps}\odot)(\hat{\delta}, \vec{g}) = \mathbf{p}(\hat{\delta}, \vec{g})\,\mathbf{s}(\hat{\delta}, \vec{b})\hat{\delta}$, where $\vec{b} := \langle g_j, g_{j+1}, \ldots\rangle$. So $\mathbf{w}(\hat{\delta}, \vec{g}) = \mathbf{p}(\hat{\delta}, \vec{g})\,\mathbf{s}(\hat{\delta}, \vec{g})\hat{\delta} = \mathbf{p}(\hat{\delta}, \vec{g})\delta = g_0\delta^t\delta = g_0\delta^{t+1}$. Furthermore, $t+1 \leqslant j \leqslant k-1$. Claim One is established.

Claim Two: Since $k \leqslant n$, for every $i < k$ there exists $\mathbf{v} \in F^\sigma(k)$ such that $\mathbf{v}(\hat{\delta}, \vec{g}) = g_0^i$.

Pick an appropriate i. Let $\mathbf{v} := x_0x_1 \bullet x_2 \bullet x_3 \bullet \cdots \bullet x_{i-1} \bullet \mathbf{r}\odot$ where $\mathbf{r} \in F^\sigma(k-i)$. Then $\mathbf{v}(\hat{\delta}, \vec{g}) =$

$$g_0g_1\hat{\delta}g_2\hat{\delta}\cdots\hat{\delta}g_{i-2}\hat{\delta}g_{i-1}\hat{\delta}\delta = g_0\delta g_2\hat{\delta}\cdots g_{i-2}\hat{\delta}g_{i-1}\hat{\delta}\delta =$$

$$g_0\delta^2g_3\hat{\delta}\cdots\hat{\delta}g_{i-1}\hat{\delta}\delta = \cdots = g_0\delta^{i-1}\delta = g_0\delta^i.$$

Claim Two is established.

If $k > n+1$ then $\delta^{k-1} = \delta^j$ for some $j \in \{1, 2, \ldots, n\}$. In the light of Claims One and Two each of the n distinct elements $g_0\delta^j \in n$ determines a distinct equivalence class $[\mathbf{w}]_{\hat{\delta}} \in F^\sigma(k)/\approx_{\hat{\delta}}$, and all of the elements in $F^\sigma(k)/\approx_{\hat{\delta}}$ are thus determined if $k > n$. □

Corollary 3.12. *If $C(k) > n$ then there exist distinct formal k–products $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{u} \approx_{\hat{\delta}} \mathbf{v}$.*

Proof. $|F^\sigma(k)| = C(k)$. Therefore if $C(k) > n$ then the pigeonhole principle applies, since $|F^\sigma(k)/\approx_{\hat{\delta}}| \leqslant n$ by Theorem 3.11. □

Conjecture 3.13. *No finite groupoid is completely free.*

Under the assumption that our conjecture is correct, it becomes relevant to raise the following question:

Problem 3.14. *Given $2 \leqslant n \in \mathbf{N}$, what is the smallest integer $\tau(n)$ such that, for every integer $k \geqslant \tau(n)$ and for every groupoid $\langle n; \diamond\rangle$, there exist elements $\mathbf{a} \neq \mathbf{b}$ in $F^\sigma(k)$ for which $\mathbf{a} \approx_\diamond \mathbf{b}$?*

4. k–Anti-Associativity

In RPN an ordered triple $\langle x, y, z\rangle$ of elements in G associates under μ iff $xy\mu z\mu = xyz\mu\mu$. RPN confers other conveniences besides relieving us of parenthesis jungles. We use it to express the complex products involving the nonassociative binary operations of concern in this section.

If the groupoid $\langle G;\mu\rangle$ happens to be a semigroup then, for every pair $\{\mathbf{p},\mathbf{s}\} \subseteq F^\sigma(k)$ of formal k–products, we get that $\mathbf{p}(\mu,\vec{g}) = \mathbf{s}(\mu,\vec{g})$ whenever $\vec{g} \in G^\infty$. That is, in a semigroup, all formal k–products are μ–equivalent. So we focus on non-semigroups. We seek groupoids which are, indeed, "as anti-associative as possible". The following remarks elaborate.

The concept of k–anti-associativity, as it pertains to a groupoid $\langle G;\mu\rangle$, is trivial for $1 \leqslant k \leqslant 2$. Henceforth we take it that $k \geqslant 3$.

$\langle G;\mu\rangle$ is 3–*anti-associative* iff $xyz\mu\mu \neq xy\mu z\mu$ for every ordered triple $\langle x, y, z\rangle \in G^3$. Theorem 3.10, and a comment preceding it on horizontal derangements, provides $2 \cdot |\mathrm{Drn}(n)|$ distinct 3–anti-associative groupoids on the set n, when $2 \leqslant n \in \mathbf{N}$.

$\langle G;\mu\rangle$ is 4–*anti-associative* iff, for every $\vec{g} := \langle g_0, g_1, g_2, g_3, \ldots\rangle \in G^\infty$, the subset,

$$\{g_0g_1g_2g_3\mu\mu\mu,\ g_0g_1g_2\mu g_3\mu\mu,\ g_0g_1\mu g_2g_3\mu\mu,\ g_0g_1g_2\mu\mu g_3\mu,\ g_0g_1\mu g_2\mu g_3\mu\},$$

of G is 5–membered. If $\langle G;\mu\rangle$ is 4–anti-associative, clearly $|G| \geqslant 5$.

Do there exist 4–anti-associative groupoids?

Definition 4.1. $\langle G;\mu\rangle$ is *k–anti-associative* iff $\mathbf{u}(\mu,\vec{g}) \neq \mathbf{v}(\mu,\vec{g})$ for all $\langle \mathbf{u}, \mathbf{v}, \vec{g}\rangle \in F^\sigma(k) \times F^\sigma(k) \times G^\infty$ with $\mathbf{u} \neq \mathbf{v}$.

Theorem 4.2. *The groupoid $\langle F^\sigma;\odot\rangle$ is k–anti-associative if $k \geqslant 3$.*

Proof. Fix $k \geqslant 3$. We normally use $\langle F^\sigma;\odot\rangle$ as a tool for evaluating the subassociativity of other groupoids. Since our argument here requires $\langle F^\sigma;\odot\rangle$ itself to be evaluated, we relabel this groupoid *qua* instrument in order to distinguish it from the same groupoid *qua* entity scrutinized.

$\langle \overline{F}^\sigma;\overline{\odot}\rangle$ is the tool version; its elements are words in the alphabet $\{\overline{\bullet}, \overline{x}_0, \overline{x}_1, \overline{x}_2, \ldots\}$. Let $\vec{\mathbf{g}} := \langle \mathbf{g}0, \mathbf{g}1, \mathbf{g}2, \ldots\rangle \in (F^\sigma)^\infty$ be any infinite sequence of finite formal products. We must prove that $\overline{\mathbf{u}}(\odot,\vec{\mathbf{g}}) \neq \overline{\mathbf{v}}(\odot,\vec{\mathbf{g}})$ for every $\{\overline{\mathbf{u}},\overline{\mathbf{v}}\} \subseteq \overline{F}^\sigma(k)$ with $\overline{\mathbf{u}} \neq \overline{\mathbf{v}}$.

For $\{\mathbf{r},\mathbf{s}\} \subseteq F^\sigma$, recall that $\mathbf{r} = \mathbf{s}$ iff $\mathbf{r}$ and $\mathbf{s}$ are spelled alike as finite words in the infinite alphabet $\{\bullet, x_0, x_1, x_2, \ldots\}$, in which event there exists $j \in \mathbf{N}$ such that $\{\mathbf{r},\mathbf{s}\} \subseteq F^\sigma(j)$.

Now choose any $\langle \overline{\mathbf{u}}, \overline{\mathbf{v}} \rangle \in \overline{F}^{\sigma}(k) \times \overline{F}^{\sigma}(k)$ such that $\overline{\mathbf{u}} \neq \overline{\mathbf{v}}$. Obviously $|\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})| = |\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})|$. So, in order to prove that $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) \neq \overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$, we must show that the words $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$ and $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$ are spelled differently.

The following remarks should be viewed in the light of Definition 2.2 and the material between that definition and the statement of Theorem 2.3.

Among the possibly many occurrences of the letter $\bullet$ in the word $\overline{\mathbf{u}}(\odot, \vec{g}) \in F^{\sigma} \subseteq \{\bullet, x_0, x_1, x_2, \ldots\}^*$ are exactly $k-1$ of them which derive from transformations of $\odot$ into $\bullet$. The same is true of the word $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}}) \in F^{\sigma}$. We tag those crucial occurrences of $\bullet$ in order to keep track of them: We write them as $\bullet'$.

If we removed all of the $k-1$ occurrences of $\bullet'$ from $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$, and all of the $k-1$ occurrences of $\bullet'$ from $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$, then the two resulting shortened words would be identical. It is the differing placements of those $k-1$ vagrant tagged $\bullet'$ in $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$ and in $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$ that make the two words differ in their spellings. We now establish this orthographic distinction.

Since by hypothesis $\overline{\mathbf{u}} \neq \overline{\mathbf{v}}$, there is a smallest integer m for which $m = |\overline{\mathbf{p}}_{\overline{\mathbf{u}}}| = |\overline{\mathbf{p}}_{\overline{\mathbf{v}}}|$, but for which $\overline{\mathbf{p}}_{\overline{\mathbf{u}}} \neq \overline{\mathbf{p}}_{\overline{\mathbf{v}}}$, where $\overline{\mathbf{p}}_{\overline{\mathbf{u}}}$ and $\overline{\mathbf{p}}_{\overline{\mathbf{v}}}$ are prefixes respectively of $\overline{\mathbf{u}}$ and $\overline{\mathbf{v}}$. Let $\mathbf{p_u}$ be the prefix that is generated in $\mathbf{u}(\odot, \vec{g})$ from $\overline{\mathbf{p}}_{\overline{\mathbf{u}}}$ under the mapping $\overline{\mathbf{u}} \mapsto \overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$. Let $\mathbf{p_v}$ be similarly obtained from $\overline{\mathbf{p}}_{\overline{\mathbf{v}}}$.

Since surely both $\overline{\mathbf{u}}$ and $\overline{\mathbf{v}}$ have $\overline{x}_0$ as a prefix, we have that $m \geqslant 2$. Furthermore, by our choice of m, if $\overline{\mathbf{q}}$ is a prefix of $\overline{\mathbf{p}}_{\overline{\mathbf{u}}}$ with $|\overline{\mathbf{q}}| = m-1$, then $\overline{\mathbf{q}}$ is a prefix also of $\overline{\mathbf{p}}_{\overline{\mathbf{v}}}$. Therefore the length-one suffix of $\overline{\mathbf{p}}_{\overline{\mathbf{u}}}$ differs from the length-one suffix of $\overline{\mathbf{p}}_{\overline{\mathbf{v}}}$.

Without loss of generality we suppose that $\overline{\mathbf{p}}_{\overline{\mathbf{u}}}$ has $\bullet$ as its length-one suffix. Then $\mathbf{p_u}$ has $\bullet'$ as its length-one suffix. Moreover, $\overline{\mathbf{p}}_{\overline{\mathbf{v}}}$ has some $\overline{x}_c$ as its length-one suffix. There are two cases.

Case: $\mathbf{g}c = x_0$. Then $|\mathbf{p_u}| = |\mathbf{p_v}|$ and $\mathbf{p_v}$ has a length-one suffix of the sort $x_d \notin \{\bullet', \bullet\}$. So the length–$|\mathbf{p_u}|$ prefix of $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$ differs from the length–$|\mathbf{p_u}|$ prefix of $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$, whence $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) \neq \overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$.

Case: $|\mathbf{g}c| \geqslant 3$, and so $|\mathbf{p_u}| < |\mathbf{p_v}|$. Then the length–$|\mathbf{p_u}|$ prefix of $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$ has some $x_d \notin \{\bullet', \bullet\}$ as its length-one suffix. So the words $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$ and $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$ have distinct length–$|\mathbf{p_u}|$ prefixes, and are therefore themselves distinct.

These cases are exhaustive, and in both cases $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) \neq \overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$. □

Illustrative Example 6: Let $\mathbf{g}0 := x_0 x_1 \bullet$ and $\mathbf{g}1 := \mathbf{g}2 := x_0$ and $\mathbf{g}3 := x_0 x_1 \bullet x_2 \bullet$ be the first four terms in a sequence $\vec{\mathbf{g}} \in (F^{\sigma})^{\infty}$ of formal

products. Consider the actions on $\vec{\mathbf{g}}$ of two elements $\overline{\mathbf{u}}$ and $\overline{\mathbf{v}}$ in the set $\overline{F}^{\sigma}(4)$; to wit, the words $\overline{\mathbf{u}} := \overline{x}_0\overline{x}_1\overline{\bullet}\,\overline{x}_2\overline{x}_3\overline{\bullet}\,\overline{\bullet}$ and $\overline{\mathbf{v}} := \overline{x}_0\overline{x}_1\overline{x}_2\overline{x}_3\overline{\bullet}\,\overline{\bullet}\,\overline{\bullet}$.

Now $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) = \mathbf{g0}\,\mathbf{g1} \odot \mathbf{g2}\,\mathbf{g3} \odot \odot = x_0x_1 \bullet x_0 \odot x_0x_0x_1 \bullet x_2 \bullet \odot\odot = x_0x_1 \bullet x_2 \bullet' x_0x_1x_2 \bullet x_3 \bullet \bullet'\odot = x_0x_1 \bullet x_2 \bullet' x_3x_4x_5 \bullet x_6 \bullet \bullet'\bullet'$, with the tags $'$ appended to those instances in the word $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}})$ of the letter $\bullet$ which came from transformed operator symbols $\odot$, which in their turn replaced the occurrences of the letter $\overline{\bullet}$ in the word $\overline{\mathbf{u}}$. In summary,

$$\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) = x_0x_1 \bullet x_2 \bullet' x_3x_4x_5 \bullet x_6 \bullet \bullet' \bullet' .$$

Likewise, $\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}}) = \mathbf{g0}\,\mathbf{g1}\,\mathbf{g2}\,\mathbf{g3} \odot\odot\odot$, and so eventually

$$\overline{\mathbf{v}}(\odot, \vec{\mathbf{g}}) = x_0x_1 \bullet x_2x_3x_4x_5 \bullet x_6 \bullet \bullet' \bullet' \bullet' .$$

Notice: $\overline{\mathbf{u}}(\odot, \vec{\mathbf{g}}) \neq \overline{\mathbf{v}}(\odot, \vec{\mathbf{g}})$ because the $\bullet'$ occur differently in each word.

Since $\langle F^{\sigma}; \odot\rangle$ achieves the theoretical extreme of anti-associativity, and since in a semigroup everything is k–associative for every k, we imagine a hierarchy of groupoids between these extremes. Of course, the set F^{σ} is infinite, rendering anti-associativity fairly straightforward to produce.

Recall that $|F^{\sigma}(k)| = C(k)$. Thus

Theorem 4.3. *If $C(k) > n$, no groupoid $\langle n; \diamond\rangle$ is k–anti-associative. So, no finite groupoid is k–anti-associative for every $k \in \mathbf{N}$.*

Problem 4.4. *For each $k \geqslant 3$ is there some $n := n(k) \in \mathbf{N}$ and some $\beta : n \times n \to n$ such that the groupoid $\langle n; \beta\rangle$ is k–anti-associative?*

Our inquiry refines and extends in natural ways. Here is one:

For integers $n \geqslant 2$ and $k \geqslant 3$ and a binary operation $\diamond : n^2 \to n$ let

$$\Psi(n, k, \diamond) := |\{\vec{g} | k : \vec{g} \in n^{\infty} \wedge \forall\{\mathbf{u}, \mathbf{v}\} \subseteq F^{\sigma}(k)\,(\,\mathbf{u}(\diamond, \vec{g}) = \mathbf{v}(\diamond, \vec{g})\,)\}|.$$

Given an arbitrary rational number $q \in [0, 1]$ does there exist a relevant triple $\langle n, k, \diamond\rangle$ such that

$$q = \frac{\Psi(n, k, \diamond)}{n^k} \, ?$$

Acknowledgments. Lício H. Bezerra and David Hobby inspired our search for Example 5. Sylvia Silberger was party to the discussion leading to this paper. Her father, Donald Silberger, is indebted to Universidade Federal de Santa Catarina, to its PET project Director, José Luiz Pinho, and to its Mathematics Chair, Nereu E. Burin, for their hospitality and moral support.

References

[1] **J. H. Folkman and R. L. Graham**: *On highly non-associative groupoids*, Colloq. Mathematicum **24** (1972), $1-10$.

[2] **H. W. Gould**: *Research Bibliography of Two Special Sequences*, Combinatorial Research Institute, West Virginia University, Morgantown, 1977.

[3] **I. M. Niven**: *Mathematics of Choice or How to Count without Counting*, final chapter. New Mathematical Library, Vol. **15**, 1965. Random House, New York.

[4] **D. M. Silberger**: *Occurrences of the integer* $(2n-2)!/n!(n-1)!$, Prace Mat. **13** (1969), $91-96$.

Universidade Federal de Santa Catarina
Florianópolis
Brazil

Received November 30, 2005

Mathematics Department
State University of New York at New Paltz
75 South Manheim Boulevard
New Paltz, NY 12561
U.S.A.
e-mail: DonSilberger@hvc.rr.com

Quasigroups and Related Systems **14** (2006), 27 – 41

The construction of loops using right division and Ward quasigroups

Kenneth W. Johnson

Abstract

Constructions are given of families of loops which can be described in terms of the table obtained from the loop by using the operation of right division. The motivation comes from group representation theory and the group matrix which goes back to Frobenius.

1. Introduction

The point of departure of this paper is the study of a group or a loop Q via the multiplication table $W(Q)$ under the operation of right division. The right division operation of a group was used by Frobenius in [3] where group characters and representation theory first appear, and the symmetry of the table corresponding to this operation is used extensively in Frobenius' work. Recently these multiplication tables were discussed in more detail in [9], where Ward quasigroups appear and some comments are made related to the extension to loops. We discuss here how interesting loops can be constructed by relaxing some of the strong symmetry of $W(G)$ for a group G. Often families of loops have been constructed by forming extensions

$$1 \to G \to Q \to H \to 1 \tag{1}$$

where G and H are groups, H is normal in Q and $Q/H \doteq G$. For example the family of Moufang loops $M(G, 2)$ constructed by Chein in [1] are of this form with an arbitrary non-commutative group G and $H \doteq C_2$, the cyclic group of order 2. A symmetric construction of $W(Q)$ arising from an extension of the form (1) can be given in this case. Several families of

2000 Mathematics Subject Classification: 20N05
Keywords: Quasigroup, Ward quasigroup, right division, Frobenius

such loops are described below. A simple example occurs when $G \doteq C_n$ and $H \doteq C_2$, and the latin square of $(Q, \backslash)$ is defined to be

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}$$

where A is $W(G)$ (which will usually be written in terms of $\{1, ..., n\}$) and B is the ordinary multiplication table of G (usually written in terms of $\{n+1, ..., 2n\}$). Then $(Q, .)$ is D_{2n}, the dihedral group of order $2n$. If we take any nonabelian group G then the corresponding loop $(Q, .)$ is non-associative and we call this the $D(G)$, the *dihedral extension* of G.

The table $W(G)$ corresponding to a group G has very strict symmetry which is explored in [9]. If some of this symmetry is relaxed, the table will correspond to a non-associative loop, but in general this loop need not have nice properties. We give examples of loops Q of order 6 which have $W(Q)$ (defined analogously) close to $W(G)$ for a group G, and which satisfy common algebraic identities. We also show how the Moufang loops $M(G, 2)$ of small orders have $W(Q)$ which can easily be described and that in particular $M(D_{2n}, 2)$ has a Ward table which is made up of block circulants and reverse circulants. In [10] a variation on a group split extension is given which produces a Bol loop. We describe the Ward tables in this case and show the tables of a group and its corresponding loop. The ideas here are in the same direction as those in the works of Drápal on distances between groups and loops [2].

In Section 2 a summary of the relevant work on Ward-Dedekind quasigroups is given, and a discussion of tables for $W(Q)$ for loops of order 6 and Moufang loops is given. The next section first describes $W(D)$ where D is a group split extension $G \times H$ arising from an action of G on H and then points out how this can be modified to produce the Bol split extension described in [10]. In Section 4 there is given the construction and some properties of dihedral extensions and Frobenius extensions and some remarks and conjectures form the remaining section.

2. Ward quasigroups and variations

Consider the quasigroup $(Q, *)$ obtained from a finite group $(G, .)$ of order n using right division which we write as $(*)$, i.e. $g * h = gh^{-1}$. It is reasonably well-known that such quasigroups may be characterised as those satisfying the identity

$$(x * z) * (y * z) = x * y.$$

Quasigroups satisfying this identity are called *Ward quasigroups*. We refer to [9] for the details. Suppose that $H =< s >$ is any cyclic subgroup of G of order m and let $\{x_1 = e, x_2, ..., x_r\}$ be a left transversal to H in G. If the left cosets of H in G are ordered as $\{eH, x_2H, ..., x_rH\}$ and each coset x_iH is ordered as

$$x_i, x_is, x_is^2, ...x_is^{m-1}$$

the multiplication table $W(Q, H)$ of $(Q, *)$ with rows and columns indexed by the elements of G in this order has the following properties.

(i) It consists of $m \times m$ blocks of circulant matrices. (The *circulant* $C(a_1, a_2, ...a_u)$ is a $u \times u$ matrix each row of which is obtained from the previous one by a right shift with wrap-around:

$$C(a_1, a_2, ...a_u) = \begin{bmatrix} a_1 & a_2 & a_3 & ... & a_u \\ a_u & a_1 & a_2 & ... & a_{u-1} \\ & & ... & & \\ a_2 & a_3 & ... & a_u & a_1 \end{bmatrix}).$$

(ii) Define the *inverse pattern* $\pi : G \to G$ by $\pi(g) = g^{-1}$. Then $W(Q, H)(i, j) = \pi(W(Q, H)(j, i))$.

(iii) $W(Q, H)(i_1, j) * W(Q, H)(i_2, j) = W(Q, H)(i_1, k) * W(Q, H)(i_2, k)$ for all i_1, i_2, j, k , i.e. the product under $*$ of corresponding elements of a fixed pair of rows is constant.

(iv) The identity element appears in the diagonal.

The following examples illustrate the symmetrical tables $W(Q, H)$ described above.

(1) Let $G = S_3$, the symmetric group on 3 objects and H be the subgroup generated by a 3-cycle. Then

$$W(Q, H) = \begin{bmatrix} 1 & 3 & 2 & 4 & 5 & 6 \\ 2 & 1 & 3 & 6 & 4 & 5 \\ 3 & 2 & 1 & 5 & 6 & 4 \\ 4 & 6 & 5 & 1 & 2 & 3 \\ 5 & 4 & 6 & 3 & 1 & 2 \\ 5 & 5 & 4 & 2 & 3 & 1 \end{bmatrix}.$$

Usually tables such as that above will be written in abbreviated form as

$$\begin{bmatrix} C(1,3,2) & C(4,5,6) \\ C(4,6,5) & C(1,2,3) \end{bmatrix} \tag{2}$$

(2) More generally if $G = D_{2n}$ and $H = C_n$ the table of $W(Q, H)$ is

$$\begin{bmatrix} C(1, n, n-1, ..., 2) & C(n+1, n+2, ..., 2n) \\ C(n+1, 2n, 2n-1, ..., n+2) & C(1, 2, ..., n) \end{bmatrix}$$

which may also be written

$$\begin{bmatrix} A^t & B \\ B^t & A \end{bmatrix}$$

where $A = C(1, 2, ..., n)$ and $B = C(n+1, n+2, ..., 2n)$.

(3) Let $G = A_4$, the alternating group. Here we take

$$H = \{e,\ (1,2)(3,4)\}$$

but we order the left cosets of H by listing the cosets of the normaliser of H, which is of course the Klein 4-group V. Specifically, the list is

$V = \{e, (1,2)(3,4), (1,4)(2,3), (1,3)(2,4)\} = H + (1,4)(2,3)H.$

$(1,2,3)V = \{(1,2,3), (2,4,3), (1,4,2), (1,3,4)\} = (1,2,3)H + (1,4,2)H$

$(1,3,2)V = \{(1,3,2), (1,4,3), (1,2,4), (2,3,4)\} = (1,3,2)H + (1,2,4)H.$

To simplify the table we write the above cosets as

$$V = \{1^1, 2^1, 3^1, 4^1\},\ (1,2,3)V = \{1^2, 2^2, 3^2, 4^2\},\ (1,3,2)V = \{1^3, 2^3, 3^3, 4^3\}.$$

Then $W(Q, H)$ is

$$\begin{bmatrix} MV(1,2,3,4)^1 & MV(1,4,3,2)^3 & MV(1,3,4,2)^2 \\ MV(1,2,3,4)^2 & MV(1,4,3,2)^1 & MV(1,3,4,2)^3 \\ MV(1,2,3,4)^3 & MV(1,4,3,2)^2 & MV(1,3,4,2)^1 \end{bmatrix}$$

where

$$MV(1,2,3,4)^i = \begin{bmatrix} 1^i & 2^i & 3^i & 4^i \\ 2^i & 1^i & 4^i & 3^i \\ 3^i & 4^i & 1^i & 2^i \\ 4^i & 3^i & 2^i & 1^i \end{bmatrix}.$$

This illustrates how the cosets of the normaliser of H can be used to further organise $W(Q, H)$.

Now consider the loops produced when some of the the restrictions on the symmetry of $W(Q, H)$ are relaxed. Given symbols $c_1, \ldots, c_n$, we denote by $R(c_1, \ldots, c_n)$ the *reversed circulant matrix*

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_2 & c_3 & \cdots & c_1 \\ \vdots & \vdots & & \vdots \\ c_n & c_1 & \ldots & c_{n-1}. \end{bmatrix}.$$

We first look at some examples of loops of order 6. If we take G as C_6 and H as C_2 then $W(Q,H)$ is

$$\begin{bmatrix} C(1,2) & C(5,6) & C(3,4) \\ C(3,4) & C(1,2) & C(5,6) \\ C(5,6) & C(3,4) & C(1,2) \end{bmatrix}.$$

There is a unique non-associative loop of order 6 which has a normal subloop of order 2 with $W(Q,H)$ as below

$$\begin{bmatrix} C(1,2) & C(5,6) & C(3,4) \\ C(3,4) & C(1,2) & C(6,5) \\ C(5,6) & C(3,4) & C(1,2) \end{bmatrix}$$

where the difference is seen to be in the $(2,3)$ block. The corresponding loop is commutative, and satisfies the weak inverse and weak inverse abelian properties.

Several loops of order 6 have a normal C_3. The closest variation on the table (2) is

$$\begin{bmatrix} C(1,3,2) & C(4,5,6) \\ C(4,6,5) & C(1,3,2) \end{bmatrix} \tag{3}$$

which corresponds to an inverse property loop. It is seen that the circulant in the $(2,2)^{\text{th}}$ place has been modified. The loop is isotopic to the loop with $W(Q,H)$

$$\begin{bmatrix} C(1,3,2) & C(4,5,6) \\ R(4,5,6) & C(1,3,2) \end{bmatrix}$$

which is commutative, and satisfies the inverse property.

Further examples are

$$\begin{bmatrix} C(1,3,2) & R(4,5,6) \\ C(4,6,5) & C(1,3,2) \end{bmatrix}$$

whose corresponding loop satisfies the right inverse property, and

$$\begin{bmatrix} C(1,3,2) & C(5,6,4) \\ R(4,5,6) & C(1,3,2) \end{bmatrix}$$

with corresponding loop satisfying the weak inverse, generalized Moufang and generalized Bol properties.

Now suppose we start with a non-associative loop G which has two-sided inverses and satisfies the automorphic inverse property. In a similar fashion

to that above the "generalized Ward quasigroup" $(Q, *)$ may be constructed from G by $x * y = xy^{-1}$. In general the left cosets of a subloop H need not partition G and even if they do the set $(a_iH)(a_jH)^{-1}$ may contain more than $|H|$ elements, but we can avoid this problem if we take H to be a normal cyclic subgroup of G. The resulting multiplication table of $(Q, *)$ satisfies (ii) and (iv) above but if G is not associative (iii) cannot be satisfied. We use $W(G, H)$ to denote the multiplication table of $(Q, *)$ with the elements of G ordered by left cosets of H exactly as above for groups.

Let P be the smallest non-associative Moufang loop M_{12} of order 12 and let H be the unique subgroup of order 3. Then $W(P, H)$ is

$$\begin{bmatrix} C(1,3,2) & C(4,5,6) & C(7,8,9) & C(10,11,12) \\ C(4,6,5) & C(1,2,3) & R(10,12,11) & R(7,9,8) \\ C(7,9,8) & R(10,12,11) & C(1,2,3) & R(4,6,5) \\ C(10,12,11) & R(7,9,8) & R(4,6,5) & C(1,2,3). \end{bmatrix} \tag{4}$$

Note that the first row and column (of blocks) and the blocks on the diagonal are determined by diassociativity and it is interesting to note how the remaining part of the table has nice symmetry. The table obviously violates condition (i) above and it is easy to see that (iii) also fails. It may also be remarked that the inverse pattern of $(Q, *)$, which may be read off from the first row of the table could not give rise to a group, as there is no group of order 12 with 9 involutions.

There are 5 nonassociative Moufang loops of order 16 and each of them has a cyclic normal subloop of order 4 (cf. [4]). The multiplication tables of the associated quasigroups can be all written in such a way that every 4×4 block in the first row, first column or along the main diagonal is a circulant, while every other block is a reversed circulant. It seems unlikely that arbitrary Moufang loops with a normal cyclic subgroup would have a table of this form, but if K is $M(D_n, 2)$ the table $T(K, C_n)$ is similar to (4) in that the blocks which are not forced by diassociativity to be circulants are reverse circulants of the form $R(m+1, m+n, m+n-1, ..., m+2)$. [1]

The Chein construction $M(G, 2)$ may be given explicitly as follows. Con-

1. The referee has informed me that there is a loop of order 32 with a normal cyclic subloop of order 4 which has off-diagonal blocks which are neither circulants nor reversed circulants.

sider the set $G \times C_2$ with multiplication

$$(g,0)(h,0) = (gh,0)$$
$$(g,0)(h,1) = (hg,1)$$
$$(g,1)(h,0) = (gh^{-1},1)$$
$$(g,1)(h,1) = (h^{-1}g,0).$$

It is well known that $M(G,2)$ is Moufang and if G is not commutative it is nonassociative. The table $W(M(G,2))$ may be written

$$\begin{bmatrix} A & B \\ B^T & \pi(A^T)\pi \end{bmatrix}$$

where A is $(W(G),0)$ and B is the table of the multiplication on the set $(G,1)$ given by $(g,1)(h,1) = (gh,1)$, π is the inverse map and we also denote by π the permutation matrix of order $n \times n$ corresponding to π.

3. Split Extensions

In [10] a variation on the split extension of two groups is used to construct a family of Bol loops. If we have two groups G and H with an homomorphism $\beta : G \to aut(H)$ the split extension $G \rtimes H$ may be constructed on the set $G \times H$ by the explicit multiplication

$$(g_1,h_1)(g_2,h_2) = (g_1g_2, h_1^{\beta(g_2)}h_2).$$

The variation on this construction to given in [10] is

$$(g_1,h_1)(g_2,h_2) = (g_1g_2, h_1^{\beta(g_2^{-1})}h_2)$$

and if $\mathrm{Im}(\beta)$ is noncommutative a non-Moufang loop $G \rtimes_B H$ which satisfies the right Bol identity

$$x(yz.y) = (xy.z)y$$

is produced.

By direct calculation the right inverse $(*)$ operations are

$$(g_1,h_1) * (g_2,h_2) = (g_1g_2^{-1}, (h_1h_2^{-1})^{\beta(g_2^{-1})})$$

in the case of the $G \rtimes H$ and

$$(g_1,h_1) * (g_2,h_2) = (g_1g_2^{-1}, (h_1h_2^{-1})^{\beta(g_2)})$$

in the case of $G \rtimes_B H$. If we suppose that $W(D)$ is produced by ordering the elements of H as $\{h_1 = e, h_2, ..., h_m\}$ according to the left cosets of a cyclic subgroup S as described above and then ordering the elements of $G \times H$ as

$$\{(g_1, h_1), (g_1, h_2), ..., (g_1, h_m), (g_2, h_1), (g_2, h_2), ..., (g_2, h_m),$$
$$..., (g_n, h_1), (g_n, h_2), ..., (g_n, h_m)\}$$

then $W(D)$ has blocks of size $m \times m$ in the $(i, j)^{\text{th}}$ position of the form

$$\{(g_i g_{j^{-1}}, W(H)(h_1^{\beta(g_j)}, ..., h_m^{\beta(g_j)})\}$$

where $(g, W(H)(t_1, ..., t_m))$ is used to denote the square obtained by replacing h_i by (g, t_i) in $W(H)$. There is a similar table for $W(B)$ where in the $(i, j)^{\text{th}}$ block there appears

$$\{(g_i g_{j^{-1}}, W(H)(h_1^{\beta(g_j^{-1})}, ..., h_m^{\beta(g_j^{-1})})\}.$$

Example. Let $G = S_3$ and $H = V_4$. Let β be the obvious isomorphism from G to $aut(H)$. We introduce the notation (g, V^κ) where $\kappa \in S_4$ to mean the table obtained from

$$W(V) = \begin{matrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{matrix}$$

by replacing the integer i by the element (g, i^κ). The group split extension is isomorphic to S_4 and one version of the $W(D)$ is

$$\begin{bmatrix} (e,V) & (\sigma^2,V^{(2,3,4)}) & (\sigma,V^{(2,4,3)}) & (\tau,V^{(2,3)}) & (\mu,V^{(2,4)}) & (\nu,V^{(3,4)}) \\ (\sigma,V) & (e,V^{(2,3,4)}) & (\sigma^2,V^{(2,4,3)}) & (\nu,V^{(2,3)}) & (\tau,V^{(2,4)}) & (\mu,V^{(3,4)}) \\ (\sigma^2,V) & (\sigma,V^{(2,3,4)}) & (e,V^{(2,4,3)}) & (\mu,V^{(2,3)}) & (\nu,V^{(2,4)}) & (\tau,V^{(3,4)}) \\ (\tau,V) & (\nu,V^{(2,3,4)}) & (\mu,V^{(2,4,3)}) & (e,V^{(2,3)}) & (\sigma,V^{(2,4)}) & (\sigma^2,V^{(3,4)}) \\ (\mu,V) & (\tau,V^{(2,3,4)}) & (\nu,V^{(2,4,3)}) & (\sigma^2,V^{(2,3)}) & (e,V^{(2,4)}) & (\sigma,V^{(3,4)}) \\ (\nu,V) & (\mu,V^{(2,3,4)}) & (\tau,V^{(2,4,3)}) & (\sigma,V^{(2,3)}) & (\sigma^2,V^{(2,4)}) & (e,V^{(3,4)}) \end{bmatrix}$$

The corresponding Bol split extension table is

$$\begin{bmatrix} (e,V) & (\sigma^2,V^{(2,4,3)}) & (\sigma,V^{(2,3,4)}) & (\tau,V^{(2,3)}) & (\mu,V^{(2,4)}) & (\nu,V^{(3,4)}) \\ (\sigma,V) & (e,V^{(2,4,3)}) & (\sigma^2,V^{(2,3,4)}) & (\nu,V^{(2,3)}) & (\tau,V^{(2,4)}) & (\mu,V^{(3,4)}) \\ (\sigma^2,V) & (\sigma,V^{(2,4,3)}) & (e,V^{(2,3,4)}) & (\mu,V^{(2,3)}) & (\nu,V^{(2,4)}) & (\tau,V^{(3,4)}) \\ (\tau,V) & (\nu,V^{(2,4,3)}) & (\mu,V^{(2,3,4)}) & (e,V^{(2,3)}) & (\sigma,V^{(2,4)}) & (\sigma^2,V^{(3,4)}) \\ (\mu,V) & (\tau,V^{(2,4,3)}) & (\nu,V^{(2,3,4)}) & (\sigma^2,V^{(2,3)}) & (e,V^{(2,4)}) & (\sigma,V^{(3,4)}) \\ (\nu,V) & (\mu,V^{(2,4,3)}) & (\tau,V^{(2,3,4)}) & (\sigma,V^{(2,3)}) & (\sigma^2,V^{(2,4)}) & (e,V^{(3,4)}) \end{bmatrix}$$

The difference is seen to be that the second members of the second and third columns are interchanged.

4. Dihedral Extensions

The author thanks the referee for the following comment. In [12] there is a classification of all constructions of loops of Bol-Moufang type with a subgroup of index 2. The dihedral extensions below appear in their notation as $(\theta_{xy}, \theta_{xy^*}, \theta_{xy^*})$.

Let G be a finite group of order n. As above define the quasigroup $(Q(G), *)$ by

$$g * h = gh^{-1}$$

for all $g, h \in G$. Associate to $Q(G)$ the Latin square $A = W(G)$ on the set $\{1, ..., n\}$ and to $(G, .)$ a Latin square B on $\{n+1, ..., 2n\}$. Suppose $\{g_1 = e, g_2, ... g_n\}$ is an ordering of G. Then $A(i, j) = k$ where $g_i * g_j = g_k$ and $B(i, j) = k + n$ where $g_i g_j = g_k$. It is easily seen that the square B is obtained from square A by operating on the columns by the inverse pattern π and then adding n to each entry. For example, if G is the cyclic group of order 3, $A = C(1, 3, 2)$ and $B = R(4, 5, 6)$. If we form the Latin square $L(G) = \begin{bmatrix} A & B \\ B & A \end{bmatrix}$ we can interpret this as $W(D(G))$ as the Ward table of the loop $D(G)$. It is easily seen that in the above example $D(G)$ is isomorphic to the dihedral group D_6 (which is also isomorphic to S_3). We call $D(G)$ the *dihedral extension* of G. If G is cyclic of order n then $D(G)$ is the usual dihedral group D_{2n}. If G is nonabelian $D(G)$ is not associative and we obtain an interesting family of loops. We can write the multiplication for $D(G)$ explicitly as follows. The elements of $D(G)$ may be taken as ordered pairs of the form (g, ϵ) where $g \in G$ and $\epsilon \in \{0, 1\}$. Then

$$\begin{aligned}
(g, 0)(h, 0) &= (gh, 0) \\
(g, 0)(h, 1) &= (gh, 1) \\
(g, 1)(h, 0) &= (gh^{-1}, 1) \\
(g, 1)(h, 1) &= (gh^{-1}, 0).
\end{aligned}$$

The non-associativity may be proved by calculating the right maps of the form $R(x, y) = R(x)R(y)R^{-1}(xy)$. Let $x = (g, \epsilon_1)$, $y = (h, \epsilon_2)$ and $z = (h, \epsilon_3)$. Then

$$xR(y, z) = \begin{cases} x & \text{if} \quad \epsilon_1 = 0 \\ (g[h, k], \epsilon_1) & \text{if} \quad \epsilon_1 = 1 \end{cases}$$

where $[h,k]$ denotes the usual commutator $h^{-1}k^{-1}hk$. It is clear that $D(G)$ is associative if and only if G is commutative.

One interesting feature of dihedral extensions is that their character theory is similar to that for groups, especially if G is taken to be of odd order. To discuss the character theory we first calculate the conjugacy classes. We calculate the map $T(x) = L(x)R^{-1}(x)$. Let $x = (g, \epsilon_1)$, $y = (h, \epsilon_2)$. Then

$$xT(y) = \begin{cases} (hgh^{-1}, \epsilon_1) & \text{if} \quad \epsilon_1 = 0, \epsilon_2 = 0 \\ (hg^{-1}h^{-1}, \epsilon_1) & \text{if} \quad \epsilon_1 = 0, \epsilon_2 = 1 \\ (hgh, \epsilon_1) & \text{if} \quad \epsilon_1 = 1, \epsilon_2 = 0 \\ (hg^{-1}h, \epsilon_1) & \text{if} \quad \epsilon_1 = 1, \epsilon_2 = 1. \end{cases}$$

Further calculation shows that for the map $L(y,z) = L(y)L(z)L(yz)^{-1}$

$$xL(y,z) = (g, \epsilon_1) \text{ or } (hgh^{-1}, \epsilon_1).$$

Since the inner mapping group of $D(G)$ is generated by $\{R(y,z), T(y), L(y,z)\}$ as y and z run through the elements of $D(G)$ it is straightforward to determine that the conjugacy class containing an element of $D(G)$ of the form $(g,0)$ is $\{(h,0); h = k^{-1}gk$ or $h = k^{-1}g^{-1}k$ for some $k \in G\}$. For an element of the form $(g,1)$ the conjugacy class is $\{(h,1); h = uk^2$, $h = u^{-1}k^2$ for $u = t^{-1}gt$ for some $t \in G$ or $h = gc$ with $c \in G'\}$. If G is of odd order the classes of $D(G)$ are easily described. The class $C_i \neq \{e\}$ of G is distinct from $C_i^* = \{g^{-1} : g \in C\}$, and gives rise to the class B_i of $D(G)$ of the form $\{(g,0) : g \in C \cup C^*\}$. In addition to the identity class there is only one other class B' which consists of all elements of the form $\{(g,1) : g \in G\}$. The characters of a loop or quasigroup are defined in [11] in terms of the association scheme arising from the classes of the quasigroup.

Theorem 1. *Let G be a group of odd order, with irreducible characters $\{\chi_1 = 1, \chi_2, \bar{\chi}_2, ...\chi_r, \bar{\chi}_r\}$. Then the characters of $D(G)$ are*

$$\{\mu_1 = 1, \mu_1', \mu_2, ..., \mu_r\}$$

where the value of $\mu_i, i > 2$ on the class $\{e\}$ is $2\chi_i$, on the class B_i is $\chi_i(C_i + C_i^*)$ and on the class B' is 0. The value of μ_1' on all classes of the form $\{g,0\}$ is 1 and on B' is -1.

Proof. It is explained in [11] that the conjugacy classes of $D(G)$ are in $1:1$ correspondence with the orbits Δ_i of $D(G) \times D(G)$ under the action

of the left and right maps of $D(G)$ and the adjacency matrices A_i of the corresponding association scheme are the incidence matrices of these orbits: if Δ_i is an orbit then $A_i(g,h) = 1$ if $(g,h) \in \Delta_i$ and 0 otherwise. Then the A_i are as follows: if C_j is the $n \times n$ incidence matrix of the class Φ_j of G the $(2n \times 2n)$ incidence matrix of the class of $D(G)$ is of the form

$$\begin{bmatrix} C_j + C_j^* & 0 \\ 0 & C_j + C_j^* \end{bmatrix}.$$

and the incidence matrix of B' is

$$\begin{bmatrix} 0 & J \\ J & 0 \end{bmatrix}$$

where J is the all 1 matrix. Now if χ is a non-trivial irreducible character of G its value on the class C_j of G is related to an eigenvalue λ of C_j by $\chi(C_j) = m_\chi \lambda / |C_j|$, where m_j is a multiplicity. It is clear that a corresponding eigenvalue of $C_j + C_j^*$ is $\lambda + \overline{\lambda}$ where $\overline{\lambda}$ is the complex conjugate of λ and which has multiplicity $2m_\chi$. This means that to each pair $\chi, \overline{\chi}$ there is a character ψ of $D(G)$which takes on the value $2\chi(e)$ on the trivial class of $D(G)$ and $\chi(C_j) + \overline{\chi(C_j)}$ on the class B_j of $D(G)$, and 0 on B'. The remaining character μ_1' necessarily takes on the value 1 on the elements of the form $(g,0)$ and -1 on elements of the form $(g,1)$. □

Example. Let G be the nonabelian Frobenius group of order 21. Its character table is

	C_0	C_1	C_1^T	C_2	C_2^T
χ_0	1	1	1	1	1
χ_1	1	1	1	ω	ω^2
$\overline{\chi_1}$	1	1	1	ω^2	ω
χ_2	3	α	$\overline{\alpha}$	0	0
$\overline{\chi_2}$	3	$\overline{\alpha}$	α	0	0

.

where $\omega = \exp(2\pi i/3)$ and $\alpha = (-1 + \sqrt{7})/2$.
The corresponding table for $D(G)$ is

	B_0	B_1	B_2	B'
μ_0	1	1	1	1
μ_1	1	1	1	-1
μ_2	2	2	-1	0
μ_3	6	-1	0	0

.

As in group character theory, the product of two characters of a loop L is defined by

$$(\chi.\psi)(x) = \chi(x)\psi(x).$$

For groups the product of characters is automatically a character since it corresponds to the tensor product of modules. This is no longer the case for loops, and leads to the coefficient ring $\mathbb{Z}(L)$ of a loop L obtained by extending $\mathbb{Z}$ by the coefficients a_i defined by

$$(\chi.\psi) = \sum a_i\chi_i(x).$$

For the dihedral extension in the case where $|G|$ is odd we show that $\mathbb{Z}(D(G)) = \mathbb{Z}$. This follows directly from the calculation

$$(\chi_i + \overline{\chi_i})(\chi_j + \overline{\chi_j}) = (\chi_i\chi_j + \overline{\chi_i\chi_j}) + (\chi_i\overline{\chi_j} + \overline{\chi_i\overline{\chi_j}}).$$

Now if $\chi_i\chi_j = \sum a_k\chi_k$ it follows that $\chi_i\chi_j + \overline{\chi_i\chi_j} = \sum a_k(\chi_k + \overline{\chi_k})$ i.e.

$$\psi_i\psi_j = \sum a_k\psi_k$$

where $i, j \geqslant 3$ and the product $\psi_i\psi_j$ where $i \leqslant 2$ is obviously ψ_k for some k.

4.1. Frobenius Extensions

Suppose we take a group G and construct the extension $Q = F(G, 2)$ on the set $G \times \{0, 1\}$ with explicit multiplication

$$(g, 0)(h, 0) = (gh, 0) = (g, 1)(h, 1)$$
$$(g, 0)(h, 1) = (hg, 1) = (g, 1)(h, 0).$$

The referee has also indicated that Frobenius extensions also appear in [12] described by the triple $(\theta_{yx}, \theta_{yx}, \theta_{xy})$. It may readily be seen that the multiplication table of $(Q, *)$ has the structure

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}$$

where A is $W(G)$ on the elements $(1, ..., n)$ and $B = (\pi W(G)^t\pi)$ on the elements $\{n+1, n+2, ...2n\}$ is the table under the operation $(g, h) \to h^{-1}g$ (here π denotes the permutation matrix corresponding to π). Such squares

occur in the work of Frobenius on group determinant factorisation [3]. It is straightforward to compute that if G is abelian then Q is the group $G \times C_2$ and that if G is nonabelian Q is a nonassociative loop that is not Moufang and unless strong conditions are placed on G it is not diassociative.

As an example we construct the table for $F(G,2)$ where $G \approx S_3$:We let

$$A = \begin{bmatrix} C(1,3,2) & C(4,5,6) \\ C(4,6,5) & C(1,2,3) \end{bmatrix}$$

and

$$B = \begin{bmatrix} C(7,9,8) & R(10,11,12) \\ R(10,11,12) & C(7,9,8) \end{bmatrix}$$

The table is then

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}.$$

We briefly summarise the character theory. Associated to each conjugacy class C of G there is (a) a conjugacy class of Q of the form $\{(g,0) : g \in C\}$ where C is a conjugacy class of G and (b) a conjugacy class of the form $\{(gG',1) : g \in C\}$, where as usual G' denoted the derived group of G. From this and using Frobenius' work we can deduce that the character table of Q consists of $2r$ linear characters where $|G/G'| = r$ and for each non-linear irreducible character of degree m of G there is a basic character of Q of degree $\sqrt{2}m$. These exhaust the characters of Q.

5. Comments and suggestions for further work

1) It appears to be a difficult problem to produce simple loops with interesting properties from methods similar to those above. However, it is certainly true that if G is a simple group with cyclic subgroup H and $W(G,H)$ is modified by rearrangement within the circulant blocks $W(Q)$ is produced for a simple loop Q. For example it may be useful to take $G = A_5$ and $H = C_5$ and search for interesting simple loops by modifying the circulant blocks.

2) The following may also be useful. Take the simple Moufang loop Q of order 120 and form $W(Q)$ by listing the cosets either of the Moufang subloop of order 12 or that of order 24.

3) Even the construction of the Moufang loop of order 12 starts with an inverse pattern (i.e. the first row of the $W(Q)$) which is different from that

of any group. It would seem to be a difficult but interesting project to find the kind of inverse patterns which can lead to Moufang or Bol loops.

4) It would seem that there should be a proof that the Moufang loops and Bol loops described in the above work satisfy the corresponding identities using the closure properties in the webs associated to the Ward tables and that such a proof may also be helpful in trying to construct new loops.

5) Non-split extensions for groups can be approached via factor sets. Perhaps one can obtain a combinatorial description of factor sets for groups and loops using the methods here (see also [8]).

6) Call a loop Q *tame* if if $W(Q)$ can be obtained from $W(G)$ for a group G by rearrangements within the circulant blocks without destroying their circulant property. One example is the loop (3) in section 2. Question: can we describe the algebraic properties of tame loops?

7) Group character theory and loop character theory consist of the information remaining in $W(Q)$ when each elements in the table is replaced by a representative of its conjugacy class. For loops it would be interesting to find some intermediate table between $W(Q)$ and this "reduced table" which could play the role of a "super character theory".

Some of the constructions here have appeared in [5], [6] and [7].

The author would like to thank the referee for helpful comments and suggestions.

References

[1] **O. Chein**: *Moufang loops of small order*, Mem. Amer. Math. Soc. **13** (1978), no. **197**.

[2] **A. Drápal**: *On minimum distances of Latin squares and the quadrangle criterion*, Acta Sci. Math. (Szeged) **70** (2004), no. **1–2**, 3 – 11.

[3] **G. Frobenius**: *Über die Primfaktoren der Gruppendeterminante*, Sber. Preuss. Akad. Wiss. Berlin (1896), 1343 – 1382 (1986) (Gesammelte Handlungen V. **3**, 38 – 77).

[4] **E. G. Goodaire, S. May and M. Raman**: *The Moufang loops of order less than* 64, Nova Science Publishers, 1999.

[5] **K. W. Johnson**: *Some recent results on quasigroup determinants*, Demonstratio Math. **24** (1991), 84 – 93.

[6] **K. W. Johnson**: *Latin square determinants II*, Discrete Math. **105** (1992), 111 – 130.

[7] **K. W. Johnson**: *Sharp characters of quasigroups*, European. J. Combinatorics **14** (1993), 103 – 112.

[8] **K. W. Johnson and C. R Leedham-Green**: *Loop cohomology*, Czech. Math. J. **40** (**115**) (1990), 182 – 194.

[9] **K. W. Johnson and P. Vojtěchovský**: *Right division in groups, Dedekind-Frobenius group matrices and Ward quasigroups*, Abh. Math. Sem. Hamburg **75** (2005), 121 – 136.

[10] **K. W. Johnson and B. Sharma**: *On a family of Bol loops*, Boll Un. Math Ital. (Algebra e Geometria Suppl.) V. **2** (1980), 119 – 126.

[11] **K. W. Johnson and J. D. H. Smith**: *Characters of finite quasigroups*, European J. Combin. **5** (1984), no. **1**, 43 – 50.

[12] **M. Kinyon, J. D. Phillips and P. Vojtěchovský**: *Loops of Bol-Moufang type with a subgroup of index* 2, Bul. Acad. Stiinte Rep. Moldova, Matematica **3** (**49**) (2005), 71 – 87.

Abington College
Penn State University
1600 Woodland Road
Abington PA 19001
U.S.A.
e-mail: kwj1@psu.edu

Received November 27, 2005

Quasigroups and Related Systems **14** (2006), 43 – 59

Geometric means and reflection quasigroups

Jimmie Lawson and Yongdo Lim

Abstract

In this paper we show how the category of reflection quasigroups forms a natural and suitable context for the development of an abstract theory of the geometric mean, as it appears in matrix and operator theory. We provide a new characterization of those quasigroups that arise when reflection quasigroups are endowed with the mean operation. We also show how the notion of the geometric mean can be enlarged to that of weighted means, develop basic properties of the latter, and illustrate their usefulness in solving equations involving the mean operation.

1. Introduction

The notion of the geometric mean of two positive real numbers, $a\#b = \sqrt{ab}$, as the solution of the equation $x^2 = ab$ can be profitably extended to much more general contexts. A natural approach in the setting of a noncommutative group G is to "symmetrize" the equation and define the geometric mean $a\#b$ of a and b to be the unique solution of the equation $xa^{-1}x = b$, provided such a unique solution exists. In the matrix group setting the equation thus assumes the form of the simplest of the matrix Riccati equations.

The Riccati equation has a natural alternative form in the setting of nonassociative algebra. Recall that the *core* of a group G is defined to be the group equipped with the binary operation $a \bullet b := ab^{-1}a$. In the core setting we are seeking a unique solution of the equation $x \bullet a = b$. The condition that this equation always have a unique solution is just the condition that $(G, \bullet)$ be a *right quasigroup* (see Section 2). Since one quickly realizes that this is frequently not the case, it is natural to look for restricted settings in which it is. Thus we are led to look among $\bullet$-closed subsets containing the

2000 Mathematics Subject Classification: 20N05
Keywords: Quasigroup, dyadic symmetric set, geometric mean, weighted mean

identity, the *twisted subgroups* (terminology courtesy of M. Aschbacher [3]), for ones that are right quasigroups. The recognition that many familiar sets of "positive objects" such as the set of positive definite symmetric or Hermitian matrices or the set of positive elements in a C^*-algebra form twisted subgroups of the multiplicative group of invertible elements suggests that this approach is worth pursuing.

At this stage it becomes most helpful to identify in an axiomatic method the type of algebraic structures that are arising as twisted subgroups of the core of a group. For a twisted subgroup A of a group G, the following properties of $(A, \bullet)$ are readily verified: for all $a, b, c \in A$,

(1) (idempotency) $a \bullet a = a$.

(2) (left symmetry) $a \bullet (a \bullet b) = b$.

(3) (left distributivity) $a \bullet (b \bullet c) = (a \bullet b) \bullet (a \bullet c)$.

These turn out to be familiar axioms. Indeed they are the first three of the four axioms of Ottmar Loos in his axiomization of symmetric spaces [14, Chapter II], the fourth one being the topological requirement that the fixed points of L_a, solutions of $a \bullet x = x$, be isolated. In the symmetric space setting the left translation $L_a(b) := a \bullet b$ represents the symmetry or point reflection of the space through the point a, so the first two axioms at least have very natural geometric motivation.

After Loos, groupoids (or magmas as some would have it) that are idempotent, left symmetric, and left distributive received rather extensive study. N. Nobusawa [15, 16] and collaborators [6], who called them symmetric sets, were among the earliest, if not the earliest. They were also investigated by R. Pierce [17, 18] under the name of "symmetric groupoids." Through the pioneering work of Joyce [5] they resurfaced in knot theory, where they are referred to as *involutory quandles*, although such structures, called "kei's," were studied as far back as 1945 by M. Takasaki [20]. (For *quandles* in general, Axiom (2) is weakened to require only that L_a be bijective; such structures are called *pseudo-symmetric sets* by Nobusawa.) The recent dissertation of D. Stanovský [19] also contains considerable information about such structures. Other references could be mentioned as well, but we return to our train of thought.

Pasting together the abstracted properties of twisted subgroups with the property of being a right quasigroup, we are led to groupoids $(X, \bullet)$ satisfying for all $a, b, c \in X$

(1) $a \bullet a = a \quad (S_a a = a)$;

(2) $a \bullet (a \bullet b) = b \quad (S_a S_a = \mathrm{id}_X)$;

(3) $a \bullet (b \bullet c) = (a \bullet b) \bullet (a \bullet c) \quad (S_a S_b = S_{S_a b} S_a)$;

(4) the equation $x \bullet a = b$ ($S_x a = b$) has a unique solution $x \in X$, called the *geometric mean*, *mean* for short, or *midpoint* of a and b, and denoted $a\#b$.

Here we follow the notation of Loos and denote by S_a (instead of L_a) left translation by a to indicate its geometric interpretation as a symmetry or point reflection. Each axiom is then given an alternative formulation in terms of these symmetries. The terminology "geometric mean" is motivated by our entire previous discussion, while the terminology "midpoint" is suggested by that fact that it is reasonable to expect that reflection through the midpoint between a and b would carry a to b.

Note that it is an immediate consequence of Axiom (2) that the left translation S_a is a bijection (with inverse S_a); hence $(X, \bullet)$ is also a left quasigroup. Thus Axioms (2) and (4) together imply that $(X, \bullet)$ is a quasigroup.

Definition 1.1. If $(X, \bullet)$ satisfies Axioms (1)-(4), then it is called a *reflection quasigroup*. The reflection quasigroups form the objects of a category for which the morphisms are $\bullet$-preserving homomorphisms.

We note that we have previously investigated reflection quasigroups in some detail in [9, 10], where we called them "dyadic symmetric sets."

We can also consider the modified category of pointed reflection quasigroups with point-preserving $\bullet$-homomorphisms for morphisms. This viewpoint allows a weakening of Axiom (4) that is often useful for verification of the axiom in specific examples.

Lemma 1.2. *Let $(X, \bullet, \varepsilon)$, $\varepsilon \in X$, satisfy Axioms (1) – (3) and for all $b \in X$, and*

(4_ε) *the equation $x \bullet \varepsilon = b$ ($S_x \varepsilon = b$) has a unique solution $x \in X$.*

Then $(X, \bullet)$ is a reflection quasigroup.

Proof. For $a, b \in X$, pick y such that $y \bullet \varepsilon = S_y \varepsilon = a$; then $S_y a = S_y S_y \varepsilon = \varepsilon$. It follows that

$$S_y x \bullet a = b \Leftrightarrow x \bullet \varepsilon = S_y S_y x \bullet S_y a = S_y b.$$

The lemma follows from hypothesis and the fact S_y is bijective. □

The lemma leads to a characterization of those twisted subgroups admitting geometric means. Note first that if A is a twisted subgroup and e is the identity, then c is a solution of $x \bullet e = b$ if and only if $ce^{-1}c = c^2 = b$. Thus the equation $x \bullet e = b$ has a unique solution if and only if every element of A has a unique square root in A, i.e., A is *uniquely* 2*-divisible*. From the preceding lemma we conclude that

Proposition 1.3. *For a twisted subgroup A of a group G, $(A, \bullet)$ is a reflection quasigroup if and only if A is uniquely* 2*-divisible.*

Example 1.4. The twisted subgroups of positive definite symmetric real matrices, positive definite Hermitian matrices, and positive elements of a C^*-algebra are all uniquely 2-divisible, a standard and well-known fact. They thus yield examples of reflection quasigroups and admit geometric means.

The geometric mean in these contexts is a known quantity in the literature and has a variety of characterizations. In a uniquely 2-divisible twisted subgroup it can be defined directly by

$$a\#b = a^{1/2}(a^{-1/2}ba^{-1/2})^{1/2}a^{1/2};$$

one sees by direct computation that this satisfies the equation $x \bullet a = xa^{-1}x = b$. There are a variety of references where the geometric mean appears (see, for example, [1, 2, 4, 7, 8, 13]). In [8] the treatment begins with the matrix Riccati equation defining the geometric mean, as we have done here.

2. The quasigroup theory of means

After our initial foray into the study of the geometric mean [8], we began to look for a suitable categorical "home" for a general theory of the geometric mean. In the preceding section we have seen that reflection quasigroups provide a natural axiomatic theory for geometric means. In this section we demonstrate that much of the basic algebraic theory of the geometric mean can be worked out in an elementary and straightforward way in the quasigroup context. We thus set forth some elementary, mostly well-known facts about quasigroups that will be pertinent and useful for our study of reflection quasigroups.

A set M equipped with a binary operation $\bullet : M \times M \to M$ is a *quasigroup* if for all $a, b \in M$, the equations $x \bullet a = b$ and $a \bullet y = b$ have unique solutions, the *right and left quotients*, usually denoted by $x = b/a$ and $y = a \setminus b$.

Remark 2.1. Alternatively one can define a quasigroup by requiring that the left and right translation maps $L_a, R_a : X \to X$ defined by $L_a(x) = a \bullet x$ and $R_a(x) = x \bullet a$ are bijections. It follows that quasigroups are cancellative.

Quasigroups form the objects of a category with corresponding morphisms those functions that are $\bullet$-homomorphisms.

We develop the quasigroup theory that pertains to reflection quasigroups in the context of general quasigroups. We assume as we go along that $(X, \bullet)$ denotes a reflection quasigroup, and point out applications of our developments to reflection quasigroups. There will be some variation in notation in the general quasigroup setting and the reflection quasigroup setting. For example, left translation is denoted L_a in the general setting and S_a in the reflection quasigroup setting. The right quotient of b by a is denoted b/a, but this same element is called the mean and denoted by $a\#b$ in the reflection quasigroup setting.

Let $(M, \bullet)$ be a quasigroup. We consider the right quotient operation $x \bullet_r y := x/y$. That is,

$$x \bullet_r y = w \iff x/y = w \iff x = w \bullet y. \tag{2.1}$$

Lemma 2.2. *Let $(M, \bullet)$ be a quasigroup.*

(i) *$(M, \bullet_r)$ is a quasigroup.*

(ii) *$(\bullet_r)_r = \bullet$.*

(iii) *A function is a $\bullet$-homomorphism if and only if it is a $\bullet_r$-homomorphism.*

(iv) *L_a distributes over $\bullet$ if and only if it distributes over $\bullet_r$.*

Proof. *(i)* The quasigroup property of $(M, \bullet_r)$ follows from (2.1):

$$x \bullet_r a = b \iff x = b \bullet a,$$

$$a \bullet_r y = b \iff a = b \bullet y \iff y = b \setminus a.$$

(ii) From

$$a = b(\bullet_r)_r c \iff a \bullet_r c = b \iff a = b \bullet c,$$

we have $(\bullet_r)_r = \bullet$.

(iii) Suppose $f : (M, \bullet) \to (Y, \bullet)$ is a homomorphism. From (2.1)

$$x \bullet_r y = w \iff x = w \bullet y \implies f(x) = f(w \bullet y) = f(w) \bullet f(y)$$
$$\iff f(x) \bullet_r f(y) = f(w).$$

The converse follows from (ii).

(iv) Immediate from (iii). □

Lemma 2.3. *Let $(M, \bullet)$ be a quasigroup.*

(i) *$(M, \bullet)$ is idempotent if and only if $(M, \bullet_r)$ is.*

(ii) *$(M, \bullet)$ is left symmetric ($L_x \circ L_x = id_M$ for all x) if and only if $(M, \bullet_r)$ is commutative.*

Proof. (i) From (2.1) $x = x \bullet x$ if and only if $x \bullet_r x = x$.

(ii) Let $a, b \in M$. There exists a unique x such that $L_x(a) = x \bullet a = b$. Assuming that L_x is involutive and applying it to the previous equation, we conclude that $a = L_x(b) = x \bullet b$. From these two equations we conclude that

$$b \bullet_r a = b/a = x = a/b = a \bullet_r b.$$

Conversely assume that $(M, \bullet_r)$ is commutative and let $x \in M$. Let $a \in M$ and set $b := x \bullet a = L_x(a)$. Then

$$x = b/a = b \bullet_r a = a \bullet_r b = a/b,$$

so $L_x(L_x(a)) = L_x(b) = x \bullet b = (a/b) \bullet b = a$. □

Remark 2.4. Consider a reflection quasigroup $(X, \bullet)$. By definition $a \# b = b/a = b \bullet_r a = a \bullet_r b$, where the last equality follows from part (ii) of the preceding lemma. Thus the operation $\#$ is commutative, equal to $\bullet_r$, and idempotent. It follows from Lemma 2.2(i) that $(X, \#)$ is also a quasigroup and from part (ii) that $(X, \#_r) = (X, \bullet)$.

Lemma 2.5. *In a quasigroup $(M, \bullet)$, the following are equivalent.*

(i) *$R^{-1}_{x \bullet y} = R^{-1}_x \bullet R^{-1}_y$, where $(R^{-1}_x \bullet R^{-1}_y)(z) = R^{-1}_x(z) \bullet R^{-1}_y(z)$.*

(ii) *If $a \bullet c = b \bullet d := m$, then $(a \bullet b) \bullet (c \bullet d) = m$.*

Proof. (i) implies (ii): Suppose that $a \bullet c = b \bullet d := m$. Then

$$a = R_c^{-1}(a \bullet c) = R_c^{-1}(m), \quad b = R_d^{-1}(b \bullet d) = R_d^{-1}(m).$$

This implies that

$$a \bullet b = R_c^{-1}(m) \bullet R_d^{-1}(m) = (R_c^{-1} \bullet R_d^{-1})(m) = R_{c \bullet d}^{-1}(m)$$

and thus $m = R_{c \bullet d}(a \bullet b) = (a \bullet b) \bullet (c \bullet d)$.

(ii) implies (i): Let x, y and m be given. Let $a = R_x^{-1}(m), b = R_y^{-1}(m)$. Then $m = a \bullet x = b \bullet y$ and then by (ii)

$$(a \bullet b) \bullet (x \bullet y) = m.$$

This implies that

$$R_{x \bullet y}^{-1}(m) = a \bullet b = R_x^{-1}(m) \bullet R_y^{-1}(m) = (R_x^{-1} \bullet R_y^{-1})(m).$$

Since m was arbitrary, we have

$$R_{x \bullet y}^{-1} = R_x^{-1} \bullet R_y^{-1}. \qquad \square$$

Remark 2.6. A binary operation $\bullet$ is called *medial* if for all a, b, c, d,

$$(a \bullet b) \bullet (c \bullet d) = (a \bullet c) \bullet (b \bullet d).$$

Observe that if the operation $\bullet$ is idempotent and $a \bullet c = m = b \bullet d$, then the right hand side of the preceding equation is also m, so the left must be also. Hence condition (ii) of Lemma 2.5, in the setting of idempotent operations, can be viewed as weakened mediality condition, called *limited mediality*. If one interprets geometric means as midpoints of some sort, then the notion of limited mediality has an intuitive geometric interpretation.

Proposition 2.7. *In a quasigroup $(M, \bullet)$, the following are equivalent.*

- (i) *$(M, \bullet)$ is left distributive, i.e., $x \bullet (y \bullet z) = (x \bullet y) \bullet (x \bullet z)$ for all x, y, z.*
- (ii) *In $(M, \bullet_r)$, for all x, y, $R_{x \bullet_r y}^{-1} = R_x^{-1} \bullet_r R_y^{-1}$.*
- (iii) *For all $a, b, c, d \in M$, if $a \bullet_r c = b \bullet_r d := m$, then $(a \bullet_r b) \bullet_r (c \bullet_r d) = m$.*

If, in addition, M is left symmetric, then $(i) - (iii)$ are equivalent to

(*iv*) *For all* $x, y, z \in M$, $(x \bullet y) \bullet z = x \bullet (y \bullet (x \bullet z))$.

Proof. The equivalence of (*ii*) and (*iii*) follows from the preceding lemma.

Suppose that (*i*) holds. We have $w = R_x^{-1}(z)$ in $(M, \bullet_r)$ if and only if $w \bullet_r x = z$ if and only if $w = z \bullet x$. Thus $R_x^{-1}(z) = z \bullet x$. From Lemma 2.2(*iv*), we have

$$R_x^{-1}(z) \bullet_r R_y^{-1}(z) = (z \bullet x) \bullet_r (z \bullet y) = z \bullet (x \bullet_r y) = R^{-1}_{x \bullet_r y}(z).$$

Thus (*ii*) holds.

Conversely assume that (*ii*) holds. Then from the previous paragraph and the hypothesis

$$z \bullet (x \bullet_r y) = R^{-1}_{x \bullet_r y}(z) = R_x^{-1}(z) \bullet_r R_y^{-1}(z) = (z \bullet x) \bullet_r (z \bullet y).$$

By Lemma 2.2(*iv*) $(M, \bullet)$ is left distributive.

The equivalence of (*i*) and (*iv*) follows by replacing z by $x \bullet z$ in each and reducing it to the other by means of left symmetry. □

We gather together the preceding results into a characterization of the mean operation # on a reflection quasigroup. The theorem follows directly from our preceding results.

Theorem 2.8. *Let* $(X, \bullet)$ *be a reflection quasigroup. Then* $(X, \#)$ *equals* $(X, \bullet_r)$ *and is a quasigroup satisfying for all* $a, b, c, d \in X$*:*

(1) (*idempotency*) $a \# a = a$;

(2) (*commutivity*) $a \# b = b \# a$;

(3) (*limited mediality*) *If* $a \# c = m = b \# d$, *then* $(a \# b) \# (c \# d) = m$.

Furthermore, $(X, \#_r) = (X, \bullet)$.

Conversely, if $(X, \#)$ *is an idempotent, commutative, limited medial quasigroup, then* $(X, \bullet) := (X, \#_r)$ *is a reflection quasigroup and* $(X, \bullet_r) = (X, \#)$.

The preceding theorem is quite satisfactory. It provides a characterization of the geometric mean operations arising from reflection quasigroups.

Remark 2.9. It was shown in [2] that the geometric mean operation of positive definite matrices satisfies the limited medial property (2.6). We have extended this result to any reflection quasigroup.

3. Dyadic powers, weighted means, and dyadic symmetric sets

Let $(X, \bullet, \varepsilon)$ be a pointed reflection quasigroup. If X is, in particular, a uniquely 2-divisible twisted subgroup, $\bullet$ is the core operation, and ε is the identity, then by the defining equation of $\varepsilon \# x$,

$$x = (\varepsilon \# x) \bullet \varepsilon = (\varepsilon \# x)\varepsilon^{-1}(\varepsilon \# x) = (\varepsilon \# x)^2$$

and thus $\varepsilon \# x = x^{1/2}$. We thus have for all $x, y \in X$,

$$\varepsilon \# x = x^{1/2} \qquad x \bullet \varepsilon = x^2 \qquad \varepsilon \bullet x = x^{-1},$$

where the last two equalities follow directly from the definition $x \bullet y = xy^{-1}x$. We take these equalities to be defining equalities for the powers $x^{1/2}$, x^2 and x^{-1} in the case of general pointed reflection quasigroups.

Note that the squaring map is right translation R_ε, the inversion map is $L_\varepsilon = S_\varepsilon$ and the square root map is left translation $L_\varepsilon^\#$ by ε in $(X, \#)$. Since

$$\varepsilon \bullet (x \bullet \varepsilon) = (\varepsilon \bullet x) \bullet (\varepsilon \bullet \varepsilon) = (\varepsilon \bullet x) \bullet \varepsilon,$$

it follows that R_ε and S_ε commute. Since $(\varepsilon \# x) \bullet \varepsilon = x$ (by the defining equation), it follows that $L_\varepsilon^\#$ given by $L_\varepsilon^\#(x) = \varepsilon \# x = x^{1/2}$ is the inverse for the bijection R_ε. Thus the square root map $L_\varepsilon^\#$ commutes with R_ε, and also commutes with S_ε since R_ε does. The mutual commutativity of these three operators allows us unambiguously to define the dyadic powers $x^{m/2^n}$ by an appropriate composition of these three maps. We note that in the uniquely 2-divisible setting, these powers agree with those computed with respect to the group operation.

Example 3.1. The additive group $\mathbb{D}$ of dyadic rationals (rational numbers with denominator a power of 2) is a uniquely 2-divisible subgroup of itself, and hence provides an important special example of a (pointed) reflection quasigroup. The reflection operation is given by $r \bullet s = 2r - s$ and the corresponding geometric mean operation is the standard midpoint operation $r \# s = (r + s)/2$. The reflection quasigroup $\mathbb{D}$ is called the *dyadic line* and a $\bullet$-homomorphism from $\mathbb{D}$ into a reflection quasigroup X is called a *dyadic geodesic.*

Theorem 3.2 (Corollary 5.8, [9]). *Let $(X, \bullet)$ be a reflection quasigroup and let $x, y \in X$. Then there exists a unique $\bullet$-homomorphism (and hence*

also #-homomorphism) γ from the dyadic line $(\mathbb{D}, \bullet)$ to X such that $\gamma(0) = x$ and $\gamma(1) = y$. For the particular case that $x = \varepsilon$ in a pointed reflection quasigroup, γ is given by $\gamma(t) = y^t$, the t^{th}-power for the dyadic t.

The proof of the theorem involves showing that $t \mapsto y^t$, as defined in the preceding remarks, is the unique $\bullet$-homomorphism sending 0 to ε and 1 to y, and then extending to the general case by pointing the reflection quasigroup at x.

Remark 3.3. A function $\beta : \mathbb{D} \to \mathbb{D}$ is a $\bullet$-homomophism if and only if it is of the form $\beta(t) = at + b$ for constants $a, b \in \mathbb{D}$. Such maps clearly preserve midpoints, hence are #-homomorphisms, and thus $\bullet$-homomorphisms. The uniqueness statement in the preceding theorem guarantees that these exhaust the $\bullet$-homomorphisms. Thus a reparametrization of a dyadic geodesic is again a dyadic geodesic if and only if the reparametrization is affine.

Definition 3.4. We define the *t-weighted mean* $x\#_t y = \gamma(t)$, where γ is the unique dyadic geodesic with $\gamma(0) = x$ and $\gamma(1) = y$.

The weighted means allow the simultaneous extension of the reflection operation and the mean operation to an all-inclusive, comprehensive setting. For a reflection quasigroup, we define

$$\Phi : \mathbb{D} \times X \times X \to X, \qquad \Phi(t, x, y) := x\#_t y.$$

We call pairs (X, Φ) arising in this way *dyadic symmetric sets*. (In our original paper [9] we defined dyadic symmetric sets to be what in this paper are called "reflection quasigroups," but the terminology there was motivated by the existence of the weighted means.)

Remark 3.5. Observe that (i) $x = x\#_0 y = \Phi(0, x, y)$, (ii) $y = x\#_1 y = \Phi(1, x, y)$, (iii) $y\#x = x\#y = x\#_{\frac{1}{2}} y = \Phi(1/2, x, y)$, (iv) $x \bullet y = x\#_{-1} y = \Phi(-1, x, y)$, and (v) $y \bullet x = x\#_2 y = \Phi(2, x, y)$. (The first two follow from the definition of the weighted mean and the others all follow from the fact that $t \mapsto x\#_t y$ is a $\bullet$- and #-homomophism from $\mathbb{D}$.) Thus for a reflection quasigroup the map Φ incorporates and extends the reflection and mean operations into a dyadic "module."

A triple (X, Φ, ε) is a *pointed dyadic symmetric set* if (X, Φ) is a dyadic symmetric set and ε is some distinguished element of X. In this case we have that $\Phi(t, \varepsilon, y) = \varepsilon\#_t y = y^t$.

Example 3.6. For a vector space V over a field with $2 \neq 0$, we have

$$\Phi : \mathbb{D} \times V \times V \to V, \qquad \Phi(t, v, w) = (1-t)v + tw.$$

(Note that $m/2^n$ is defined in the field for all integers m, n.)

We list elementary properties of a dyadic symmetric set.

Proposition 3.7. *Let (X, Φ, ε) be a pointed dyadic symmetric set. Then the following properties hold for general elements of $\mathbb{D}$ and X:*

(*i*) *The map $\Phi_{x,y} : \mathbb{D} \to X$ defined by $\Phi_{x,y}(t) = x\#_t y$ is a $\bullet$- and $\#$-homomorphism; in particular, this holds for $t \mapsto x^t : \mathbb{D} \to X$.*

(*ii*) $x\#_t(x\#_s y) = x\#_{ts} y$; *in particular,* $(x^s)^t = x^{st}$.

(*iii*) $x\#_t y = y\#_{1-t} x$.

(*iv*) $x \bullet (y\#_t z) = (x \bullet y)\#_t(x \bullet z)$; *in particular,* $(y\#_t z)^{-1} = y^{-1}\#_t z^{-1}$.

(*v*) *The maps $x \mapsto x\#_t b$, $t \neq 1$, and $y \mapsto a\#_t y$, $t \neq 0$, are bijective.*

Proof. Part (i) follows from the definition of $x\#_t y$. Parts $(ii) - (iv)$ follow from Theorem 3.2 (the uniqueness of the defining homomorphism) and the observation that affine maps $f(t) = at + b$ are $\bullet$-homomorphisms on $\mathbb{D}$, and isomorphisms for $a \neq 0$. For (ii) the two sides of the equality, thought of as maps of t, are $\bullet$-homomorphisms (the right hand side involves composition with $f(t) = st$) sending 0 to x and 1 to $x\#_s y$, and hence agree everywhere. For part (iii) the homomorphisms again agree at 0 and 1. For (iv) the left hand composition with S_x is a $\bullet$-homomorphism as a consequence of the distributive law and the two sides again agree at 0 and 1 (the second part is the special case $x = \varepsilon$). For (v), we may assume, by changing the pointing if necessary, that $a = \varepsilon$. Then $a\#_t y = \varepsilon\#_t y = y^t$. Since we saw at the beginning of the section that the map $y \mapsto y^t$, $t \neq 0$, is some appropropriate composition of the bijections S_ε, L_ε and $L_\varepsilon^\#$, we conclude that $y \mapsto y^t = a\#_t y$ is bijective. The bijectivity of $x \mapsto x\#_t b$ then follows from (iii). □

Lemma 3.8. *Let $(X, \bullet)$ be a reflection quasigroup, and let $m \in X$. Then $X_m := \{(a, b) \in X \times X | a\#b = m\}$ is a subquasigroup of the product quasigroup $X \times X$.*

Proof. Let $(a,b),(c,d) \in X$, i.e., $a\#b = m = c\#d$. By the defining equation of $\#$, $m \bullet a = (a\#b) \bullet a = b$ and similarly $m \bullet c = d$. Thus

$$m \bullet (a \bullet c) = (m \bullet a) \bullet (m \bullet c) = b \bullet d,$$

and hence $m = (a \bullet c)\#(b \bullet d)$. It follows from the limited medial property that $m = (a\#c)\#(b\#d)$. □

The preceding lemma allows us to derive an extended version of limited mediality in the context of dyadic symmetric sets.

Corollary 3.9. *Let $(X,\bullet)$ be a reflection quasigroup and suppose that $a\#c = m = b\#d$. Then $(a\#_t b)\#(c\#_t d) = m$ for all t.*

Proof. We note that $(a,c),(b,d) \in X_m$ as defined in Lemma 3.8. Since by that lemma X_m is a subquasigroup of $X \times X$ and hence a reflection quasigroup, there exists a dyadic geodesic $t \mapsto (a,c)\#_t(b,d)$ in X_m. Since the operations are defined coordinatewise, this map is a $\bullet$-homomorphism in each coordinate, and by uniqueness of this homomorphism we must have $(a,c)\#_t(b,d) = (a\#_t b, c\#_t d)$ for each t. Since the image is in X_m for each m, the corollary follows. □

Recall from Theorem 2.8 that the quasigroup $(X,\#)$ derived from a reflection quasigroup can be characterized as a quasigroup satisfying

(1) (idempotency) $a\#a = a$;

(2) (commutativity) $a\#b = b\#a$;

(3) (limited mediality) If $a\#c = m = b\#d$, then $(a\#b)\#(c\#d) = m$.

In addition we have seen that the map $\gamma(t) = a\#_t b$ is a $\#$-homomorphism from $\mathbb{D}$ to X carrying 0 to a and 1 to b. Thus

(4) $(a\#_r b)\#(a\#_s b) = \gamma(r)\#\gamma(s) = \gamma((r+s)/2) = a\#_{(r+s)/2} b$.

We can obtain a version of Theorem 2.8 for dyadic symmetric sets by showing that all of these properties generalize to the setting of weighted means and that these generalized properties characterize dyadic symmetric sets. The preceding properties are obtained from the corresponding ones in the following theorem by specializing to the case $t = 1/2$.

Theorem 3.10. *In a reflection quasigroup $(X,\bullet)$ the weighted means satisfy the following properties for all $a,b,c,d \in X$, all $r,s,t \in \mathbb{D}$:*

(0) $a\#_0 b = a$, $a\#_1 b = b$;

(1) (*idempotency*) $a\#_t a = a$;

(2) (*commutivity*) $a\#_t b = b\#_{1-t} a$;

(3) (*limited mediality*) *If* $a\#c = m = b\#d$, *then* $(a\#_t b)\#(c\#_t d) = m$;

(4) (*affine change of parameter*) $(a\#_r b)\#_t(a\#_s b) = a\#_{(1-t)r+ts} b$;

(5) (*exponential law*) $a\#_r(a\#_s b) = a\#_{rs} b$;

(6) (*cancellativity*) $a\#_t b = a\#_t c$ *for* $t \neq 0$ *implies* $b = c$.

Conversely if for $\Phi : \mathbb{D} \times X \times X \to X$, $a\#_t b := \Phi(t, a, b)$ *satisfies items* $(0) - (6)$, *then* $a\#_t b$ *is the* t*-weighted mean for the reflection quasigroup* X *with operations* $a \bullet b := a\#_{-1} b$ *and* $a\# b := a\#_{1/2} b$.

Proof. Property (0) holds by Remark 3.5. The unique dyadic geodesic carrying 0 and 1 to a is the constant map to a, so (1) is satisfied. Properties (2) and (5) we have already established in Proposition 3.7 and property (3) in Corollary 3.9. For (4) the left hand side is a $\bullet$-homomorphism in t, and the right hand side is also, since it is the composition of the dyadic geodesic $t \mapsto a\#_t b$ with the affine map on $\mathbb{D}$ sending t to $(1-t)r + ts$ (see Remark 3.3). Since they both sent 0 to $a\#_r b$ and 1 to $a\#_s b$, by uniqueness they agree. Property (6) follows from Proposition 3.7.

Conversely suppose that items (1) through (6) are satisfied. We set $a\# b := a\#_{1/2} b$. As we remarked before the theorem, properties $(1) - (3)$ ensure that the corresponding properties of Theorem 2.8 are satisfied by $\#$. Note that the equation $a\#x = b$ has solution $x = a\#_2 b$ since $a\#_{1/2}(a\#_2 b) = a\#_1 b = b$ by Properties (5) and (0). The uniqueness follows from (6), and then by commutivity $(X, \#)$ is a quasigroup. Thus by Theorem 2.8, $(X, \bullet) := (X, \#_r)$ is a reflection quasigroup with $\#$ as its mean. From

$$(a\#_2 b)\# a = a\#_{1/2}(a\#_2 b) = a\#_1 b = b,$$

we conclude that $b\#_r a = a\#_2 b = b\#_{-1} a$. Thus $a \bullet b = a\#_{-1} b$. Finally it follows from Property (4) with $t = 1/2$ that $t \mapsto a\#_t b$ is a $\#$-homomorphism, and hence a $\bullet$-homomorphism. Thus it is the unique dyadic geodesic carrying 0 to a and b to 1. $\square$

4. Solving equations involving means

In this section we apply the machinery of weighted means that we have developed to the solution of equations involving the mean operation. There is no attempt here to develop a comprehensive theory–only to illustrate how the need for such machinery arises and how it can be employed. In [12] we have studied lower degree symmetric matrix equations in some detail. There again the mean was a crucial tool, although it did not appear directly in the equations in the setting.

Throughout this section we assume that $(X, \bullet)$ is a dyadic symmetric set and $a\#b$ is the associated geometric mean operation on X. *We also assume that the weighted mean extends to all real numbers t and that the properties of the weighted mean developed in the last section remain valid in this context.* Many of the typical examples (positive definite matrices, positive elements of a C^*-algebra), indeed most topological examples, satisfy this requirement. We have studied in some detail the topological setting for the theory of dyadic symmetric sets and the extension of the weighted mean to all real parameters in [11].

Theorem 4.1. *The geometric mean $x = a\#b$ is the unique solution of*

$$(x\#a)\#(x\#b) = x. \tag{4.2}$$

Proof. Let $x = a\#b$. Then $x = x\#x = a\#b = (x\#a)\#(x\#b)$ by the limited medial property and hence $a\#b$ is a solution of the equation (4.2).

Conversely, suppose that $(x\#a)\#(x\#b) = x$. Then

$$x\#b = x \bullet (x\#a) = (x \bullet x)\#(x \bullet a) = x\#(x \bullet a)$$

and by the cancellative law $b = x \bullet a$. Therefore, $x = a\#b$. □

Theorem 4.2. *The weighted mean $a\#_{\frac{4}{3}}b$ is the unique solution of the equation*

$$x\#(x\#a) = b.$$

Furthermore, $a\#_{\frac{2}{3}}b$ is the unique solution of the equation

$$(a\#x)\#b = x.$$

Proof. Applying the left translation $a\#_{4/3}(\cdot)$ to

$$b = x\#(x\#a) = x\#_{1/2}(x\#_{1/2}a) = x\#_{1/4}a = a\#_{3/4}x,$$

we obtain $a\#_{4/3}b = a\#_{4/3}(a\#_{3/4}x) = x$. Conversely if $x = a\#_{4/3}b$, then

$$x\#(x\#a) = x\#_{1/4}a = a\#_{3/4}x = a\#_{3/4}(a\#_{4/3}b) = a\#_1 b = b.$$

Next, consider the equation $(a\#x)\#b = b\#(a\#x) = x$. By the defining property of the mean, this is equivalent to

$$b = x \bullet (a\#x) = (x \bullet a)\#(x \bullet x) = (x \bullet a)\#x. \tag{4.3}$$

Setting $y = x \bullet a$, we have $x = y\#a$ and thus (4.3) becomes

$$y\#(y\#a) = b.$$

This has the unique solution $y = a\#_{\frac{4}{3}}b$ and hence

$$x = y\#a = (a\#_{4/3}b)\#a = a\#_{1/2}(a\#_{4/3}b) = a\#_{\frac{2}{3}}b. \qquad \square$$

Remark 4.3. We observe that for a reflection quasigroup $(X, \bullet)$ possessing general weighted means, the two quasigroups $(X, \bullet)$ and $(X, \#)$ are *orthogonal* in the sense that given $a, b \in X$, the simultaneous equations

$$\begin{cases} a &= x \bullet y \\ b &= x\#y \end{cases}$$

have a unique solution. Indeed since $a = x \bullet y$ is equivalent to $x = a\#y$, the second equation reduces to

$$b = y\#(y\#a).$$

By Theorem 4.2, we can uniquely solve the system by

$$y = a\#_{\frac{4}{3}}b, \quad x = a\#(a\#_{\frac{4}{3}}b) = a\#_{\frac{2}{3}}b.$$

Theorem 4.4. *Let $(X, \bullet, \varepsilon)$ be a pointed dyadic symmetric set and $a, b \in X$. Then the simultaneous equations*

$$\begin{cases} c \bullet a &= b\#x \\ c \bullet x &= a\#x \end{cases}$$

have unique solutions in x and c, namely $x = a\#b$ and $c = a\#(b\#(a\#b)) = (a\#b)\#(a\#(a\#b)) = a\#_{3/8}b$. In particular, the geometric mean $x = a\#b$ is the unique solution of

$$a\#(b\#x) = x\#(x\#a).$$

Proof. The second equation is equivalent to

$$x = c \bullet (c \bullet x) = c \bullet (a \# x) = (c \bullet a) \# (c \bullet x).$$

By the defining equation of the geometric mean, this is equivalent to

$$c \bullet a = x \bullet (c \bullet x).$$

Combining this with the two equations of the theorem, we have

$$b \# x = c \bullet a = x \bullet (c \bullet x) = x \bullet (a \# x) = (x \bullet a) \# (x \bullet x) = (x \bullet a) \# x.$$

From the cancellative property of the mean operation, $b = x \bullet a$ and therefore $x = a \# b$ by the mean defining equation. The characterizations for c in order that the original equations be satisfied for $x = a \# b$ then follow directly from those equations. By properties of the weighted mean in the previous section, these both reduce to $a \#_{3/8} b$, and hence are equal. Their equality ensures that $a \# b$ is a solution of the equation in the last assertion of the theorem.

Suppose that $a \# (b \# x) = x \# (x \# a)$. Setting $c = b \# x$ and $d = x \# a$, we have from the limited medial property that

$$a \# c = x \# d = (a \# x) \# (c \# d) = d \# (c \# d).$$

From the cancellation property, $x = c \# d = (x \# b) \# (x \# a)$. By Theorem 4.1, $x = a \# b$. Thus the solution in the last equation of the theorem is unique. □

References

[1] **T. Ando**: *Topics on Operator Inequalities*, Lecture Notes Hokkaido Univ., Sapporo, 1978.

[2] **T. Ando, C.-K. Li, R. Mathias**: *The geometric mean*, Linear Algebra Appl. **385** (2004), 305 – 334.

[3] **M. Aschbacher**: *Near subgroups of finite groups*, J. Group Theory **1** (1998), 113 – 129.

[4] **M. Fiedler and V. Pták**: *A new positive definite geometric mean of two positive definite matrices*, Linear Algebra Appl. **251** (1997), 1 – 20.

[5] **D. Joyce**: *A classifying invariant of knots, the knot quandle*, J. Pure Appl. Alg. **23** (1982), 37 – 66.

[6] **M. Kano, H. Nagao, and N. Nobusawa**: *On finite homogeneous symmetric sets*, Osaka J. Math. **13** (1976), 399 – 406.

[7] **F. Kubo and T. Ando**: *Means of positive linear operators*, Math. Ann. **246** (1980), 205 – 224.

[8] **J. D. Lawson and Y. Lim**: *The geometric mean, matrices, metrics, and more*, Amer. Math. Monthly **108** (2001), 797 – 812.

[9] **J. D. Lawson and Y. Lim**: *Symmetric sets with midpoints and algebraically equivalent theories*, Results in Mathematics **46** (2004), 37 – 56.

[10] **J. D. Lawson and Y. Lim**: *Means on dyadic symmetric sets and polar decompositions,* Abh. Math. Sem. Univ. Hamburg **74** (2004), 135 – 150.

[11] **J. D. Lawson and Y. Lim**: *Symmetric spaces with convex metrics*, to appear Forum Math.

[12] **J. D. Lawson and Y. Lim**: *Solving symmetric matrix word equations via symmetric space machinary*, Linear Algebra and Appl. **414** (2006), 560 – 569.

[13] **Y. Lim**: *Geometric means on symmetric cones*, Arch. der Math. **75** (2000), 39 – 45.

[14] **O. Loos**: Symmetric spaces, I: General Theory, Benjamin, New York, Amsterdam, 1969.

[15] **N. Nobusawa**: *On symmetric structure of a finite set*, Osaka J. Math. **11** (1974), 569 – 575.

[16] **N. Nobusawa**: *Orthogonal groups and symmetric sets*, Osaka J. Math. **20** (1983), 5 – 8.

[17] **R. Pierce**: *Symmetric groupoids*, Osaka J. Math. **15** (1978), 51 – 76.

[18] **R. Pierce**: *Symmetric groupoids II*, Osaka J. Math. **16** (1979), 317 – 348.

[19] **D. Stanovský**: *Left distributive left quasigroups*, Ph.D. thesis, Charles University, Prague, 2004.

[20] **M. Takasaki**: *Abstraction of symmetric transformations* (in Japanese), Tohaku Math. J. **49** (1942/3), 145 – 207.

Received November 6, 2005

Jimmie Lawson
Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803
U.S.A.
e-mail: lawson@math.lsu.edu

Yongdo Lim
Department of Mathematics
Kyungpook National Universit
Taegu 702-701
Korea
e-mail: ylim@knu.ac.kr

Quasigroups and Related Systems **14** (2006), 61 – 72

Ranks of nets

G. Eric Moorhouse

Abstract

Let $\mathcal{N}$ be a k-net of prime order p. We find bounds on the p-rank of (the point-line incidence matrix of) $\mathcal{N}$ for $k \in \{3,4\}$, and observe connections between the p-rank and certain structural properties of $\mathcal{N}$. Implications for the study of finite projective planes are described.

1. Loops and 3-nets of prime order

Let $(L,*)$ be a loop of prime order p. The 3-*net* $\mathcal{N} = \mathcal{N}(L)$ coordinatized by L is the point-line incidence system having p^2 points $L^2 = L \times L$, and $3p$ lines given by

$$\{a\} \times L \text{ for } a \in L \text{ (the lines "}x = a\text{")};$$

$$L \times \{b\} \text{ for } b \in L \text{ (the lines "}y = b\text{")}; \text{ and}$$

$$\{(x,y) \in L^2 : x * y = c\} \text{ for } c \in L \text{ (the lines "}x * y = c\text{")}.$$

The point-line incidence matrix of $\mathcal{N}$ is the $p^2 \times 3p$ matrix with rows and columns indexed by points and lines of $\mathcal{N}$ respectively; and having entries 0 and 1 corresponding to non-incident and incident point-line pairs respectively. We have

Theorem 1.1. (Main Theorem [5]) *The p-rank of the incidence matrix of $\mathcal{N}$ equals $3p-3$ if L is associative, and $3p-2$ otherwise.*

Our original proof [5], still the simplest proof available, uses loop theory. (Here for simplicity we consider only loops and nets of prime order, although

2000 Mathematics Subject Classification: 51A05, 20N05
Keywords: net, projective geometry, loop, web, p-rank

more arbitrary finite orders were considered in [5].) We reproduce this proof below; and we indicate three alternative proofs of the same result. Our (currently unrealized) goal is a generalization of Theorem 1.1 to k-nets for $k = 3, 4, \ldots, p+1$; possibly using techniques from nonassociative algebra, or possibly by generalizing some of the other techniques described in this paper. The desired generalization of this result is

Conjecture 1.2. [5] *Let $\mathcal{N}$ be any k-net of prime order p, and let $\mathcal{N}'$ be any $(k-1)$-subnet of $\mathcal{N}$ obtained by deleting one of the k parallel classes of lines of $\mathcal{N}$; here $k \in \{2, 3, \ldots, p+1\}$. Then the p-rank of the incidence matrix of $\mathcal{N}$ exceeds that of $\mathcal{N}'$ by at least $p-k+1$.*

The significance of Conjecture 1.2 lies in the fact [5] that this would imply that every projective plane of prime order is Desarguesian, thereby settling one of the most celebrated currently open problems in finite geometry. Extensions of this method to other finite orders would yield restrictions on the possible orders of finite projective planes, beyond the restrictions available through the Bruck-Ryser Theorem [2]. We believe that these finite geometric questions are worthy of the attention of researchers in nonassociative algebra. Indeed, Belousov [1] attributes the origins of quasigroup theory to the study of finite projective planes. (I am grateful to V.V. Goldberg for bringing this reference to my attention during our Mile High Conference.)

In Section 2, we describe the p-rank of a net in terms recognizable to researchers of webs. This leads to a reformulation of our main result Theorem 1.1 in equivalent terms as Theorem 2.3. In Sections 3, 4, 5 and 6 we provide proofs of this main result using loop theory, group theory, finite field theory, and number theory (specifically, exponential sums) respectively. Each of these approaches suggests different possibilities for generalization to k-nets. Finally in Section 7 we describe some recent progress towards Conjecture 1.2 in the case of 4-nets.

2. Nets and planes of prime order

Consider a field $F = \mathbb{F}_p$ of prime order p, and let $k \geqslant 2$. For every $J \subseteq \{1, 2, \ldots, k\}$ we consider the projection $F^k \to F^{|J|}$ defined by

$$(a_1, a_2, \ldots, a_k) \mapsto (a_j : j \in J).$$

We simply write $\pi_i = \pi_{\{i\}}$, $\pi_{ij} = \pi_{\{i,j\}}$, and we denote $J' = \{1, 2, \ldots, k\} \setminus J$ so that in particular

$$\pi_{i'}(a_1, a_2, \ldots, a_k) = (a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k).$$

A *k-net of order p* is a subset $\mathcal{N} \subseteq F^k$ such that for all $i \neq j$ in $\{1, 2, \ldots, k\}$, the map $\mathcal{N} \xrightarrow{\pi_{ij}} F^2$ is bijective. The members of $\mathcal{N}$ are called *points,* and the *lines* of $\mathcal{N}$ are the fibres

$$\mathcal{N} \cap \pi_i^{-1}(a) = \{v \in \mathcal{N} : \pi_i(v) = a\}$$

for $i \in \{1, 2, \ldots, k\}, a \in F$. For every $J \subseteq \{1, 2, \ldots, k\}$ of cardinality at least 2, clearly $\pi_J(\mathcal{N})$ is a $|J|$-net of order p; we call this a $|J|$-*subnet* of $\mathcal{N}$. In particular for each $i \in \{1, 2, \ldots, k\}$, we have that $\pi_{i'}(\mathcal{N})$ is a $(k-1)$-subnet of $\mathcal{N}$, obtained by simply deleting from $\mathcal{N}$ the i-th parallel class of lines. An *isomorphism* of nets $\phi : \mathcal{N} \to \mathcal{N}'$ is a map of the form $(a_1, a_2, \ldots, a_k) \mapsto (\alpha_1(a_{\sigma(1)}), \alpha_2(a_{\sigma(2)}), \ldots, \alpha_k(a_{\sigma(k)}))$ for some $\alpha_1, \alpha_2, \ldots, \alpha_k \in Sym(F)$ and $\sigma \in S_k$; this simply says that the corresponding point-line incidence structures are isomorphic.

An *affine plane* of order p is simply a $(p+1)$-net of order p. The *Desarguesian affine plane* is the $(p+1)$-net

$$\mathcal{D} = \{(a, b, a+b, a+2b, \ldots, a+(p-1)b) : a, b \in F\}.$$

A *Desarguesian net* is any subnet of $\mathcal{D}$. A Desarguesian 3-net is known simply as a *cyclic 3-net.* Every cyclic 3-net of order p is isomorphic to $\{(a, b, a+b) : a, b \in F\}$.

Denote by $\mathcal{V} = \mathcal{V}(\mathcal{N})$ the vector space consisting of all k-tuples $(f_1, f_2, \ldots, f_k)$ of functions $F \to F$ such that

$$f_1(a_1) + f_2(a_2) + \cdots + f_k(a_k) = 0$$

for all $(a_1, a_2, \ldots, a_k) \in \mathcal{N}$. Also denote by $\mathcal{V}_0 = \mathcal{V}_0(\mathcal{N}) \leqslant \mathcal{V}$ the subspace consisting of all $(f_1, f_2, \ldots, f_k) \in \mathcal{V}$ satisfying the additional condition $f_1(0) = f_2(0) = \cdots = f_k(0) = 0$. The map $\mathcal{V} \to F^k$, $(f_1, f_2, \ldots, f_k) \mapsto (f_1(0), f_2(0), \ldots, f_k(0))$ induces an isomorphism from $\mathcal{V}/\mathcal{V}_0$ to a $(k-1)$-dimensional subspace of F^k; thus $\dim(\mathcal{V}) = \dim(\mathcal{V}_0) + k - 1$, and so we may focus our attention on $\mathcal{V}_0$ rather than on $\mathcal{V}$ itself. Since $\mathcal{V}$ may be interpreted as the right null space of the point-line incidence matrix A of $\mathcal{N}$, this gives

Proposition 2.1. *The p-rank of the incidence matrix A of $\mathcal{N}$ is given by*

$$\mathrm{rank}_p A = pk - \dim \mathcal{V} = (p-1)k + 1 - \dim \mathcal{V}_0.$$

Rephrasing our conjectured bounds for the rank of A in terms of the nullity gives

Conjecture 2.2. *We have:*

(*i*) $\dim \pi_1(\mathcal{V}) \leqslant k-1$.

(*ii*) $\dim(\mathcal{V}_0) \leqslant \frac{1}{2}(k-1)(k-2)$, *and equality holds iff $\mathcal{N}$ is Desarguesian.*

Statement (*i*) is equivalent to Conjecture 1.2, and the first assertion of (*ii*) is implied by (*i*). If either (*i*) or (*ii*) holds then every plane of prime order is Desarguesian. As indication that $\mathcal{V}_0$ is more natural to consider than the row or column space of A itself, we observe that in the case of webs, the corresponding incidence map has infinite rank, whereas the null space $\mathcal{V}$ is finite-dimensional. Indeed the bound $\dim(\mathcal{V}_0) \leqslant \frac{1}{2}(k-1)(k-2)$ holds for k-webs, with equality attainable in the case of algebraic webs; see [3,4]. We rephrase the Main Theorem as

Theorem 2.3. *Let $\mathcal{N}$ be a 3-net of order p. Then* $\dim(\mathcal{V}_0) \leqslant 1$. *Moreover, equality holds iff $\mathcal{N}$ is cyclic, in which case $\mathcal{V}_0$ is spanned by a triple (f,g,h) in which the maps $f,g,h : F \to F$ are permutations.*

3. First proof of main theorem (using loop theory)

Let $\mathcal{N} \subset F^3$ be a 3-net of prime order p, in the notation of Section 2, and suppose $(f,g,h) \in \mathcal{V}_0(\mathcal{N})$ is nonzero. To within an isomorphism of nets, we have

$$\mathcal{N} = \{(x, y, x * y) : x, y \in F\}$$

where $(x,y) \mapsto x * y \in F$ is a loop operation on F with identity 0. By definition we have

$$\begin{aligned} &f(0) = g(0) = h(0) = 0; \\ &f(x) + g(y) + h(x * y) = 0 \quad \text{for all } x, y \in F. \end{aligned}$$

This implies that $f(x) = g(x) = -h(x)$ for all $x \in F$ and that f is a nonzero homomorphism from the loop $(F, *)$ to the cyclic group $(F, +)$ of order p. These two loops are therefore isomorphic, so $\mathcal{N}$ is cyclic. Moreover

every such homomorphism has the form cf for some $c \in F$, so $\mathcal{V}_0(\mathcal{N})$ is 1-dimensional. The result follows.

The same argument actually yields the stronger result

Theorem 3.1. [5] *Let L be a loop of order $n = p^r m$ where $gcd(p, m) = 1$. Then the p-rank of the incidence matrix of the 3-net $\mathcal{N}(L)$ equals $3p - 2 - e$ where $e \in \{0, 1, 2, \ldots, r\}$. We have $|K| = p^e$ where $K \subseteq L$ is the largest normal subloop such that the quotient loop L/K is an elementary abelian p-group.*

4. Second proof of main theorem (using permutation groups)

An alternative proof of Theorem 3.1 is obtained by considering the left multiplication group of L. More generally, let Ω be a set of size $|\Omega| = n = p^r m$ where $gcd(p, m) = 1$, and let G be a group permuting Ω transitively. Let $H \leqslant G$ be the stabilizer of a point which we denote $1 \in \Omega$. For each $k \geqslant 0$, denote by C^k the vector space over F consisting of all functions $\Omega^{k+1} \to F$. Then G acts on C^k via

$$f^g(x_0, x_1, \ldots, x_k) = f(x_0^g, x_1^g, \ldots, x_k^g)$$

for $g \in G$, $f \in C^k$, $x_i \in \Omega$. Consider the F-linear map $\partial = \partial_k : C^k \to C^{k+1}$ defined by

$$(\partial f)(x_0, x_1, \ldots, x_{k+1}) = \sum_{i=0}^{k+1} (-1)^{k+1-i} f(x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{k+1})$$

for $f \in C^k$, $x_i \in \Omega$. Note that ∂ is G-equivariant: $\partial(f^g) = (\partial f)^g$. The image $B^1 = \partial C^0 \leqslant C^1$ consists of all functions $\partial\phi(x_0, x_1) = \phi(x_0) - \phi(x_1)$ for some $\phi : \Omega \to F$. Consider the subspace of G-invariants given by

$$(B^1)^G = \{f \in B^1 : f^g = f \text{ for all } g \in G\}.$$

In the following, $Hom(G/K, F)$ denotes the vector space over F consisting of homomorphisms from the multiplicative group G/K to the additive group of F.

Lemma 4.1. $(B^1)^G \cong Hom(G/K, F)$ *where K is the smallest normal subgroup of G containing H such that G/K is an elementary abelian p-group. In particular,* $\dim (B^1)^G = e \in \{0, 1, 2, \ldots, r\}$ *where $|G/K| = p^e$.*

Proof. For each $\phi : G/K \to F$, define $\widehat{\phi} : \Omega \to F$ by $\widehat{\phi}(1^g) = \phi(Kg)$ for $g \in G$. Note that $\widehat{\phi} \in C^0$ is well-defined since K contains H. We claim that the map

$$\theta : Hom(G/K, F) \to (B^1)^G, \quad \phi \mapsto \partial\widehat{\phi}$$

is an isomorphism of vector spaces over F. Certainly if $\phi \in Hom(G/K, F)$ then

$$\begin{aligned}\partial\widehat{\phi}(1^{ug}, 1^{vg}) &= \phi(Kug) - \phi(Kvg) = \phi(Ku) + \phi(Kg) - \phi(Kv) - \phi(Kg) \\ &= \phi(Ku) - \phi(Kv) = \partial\widehat{\phi}(1^u, 1^v)\end{aligned}$$

for all $u, v, g \in G$. Since G is transitive on Ω, this implies that $\partial\widehat{\phi} \in (B^1)^G$. If $\partial\widehat{\phi} = 0$ then $\phi(Kg) = \phi(K) = 0$ for all $g \in G$, i.e. $\phi = 0$ so θ is injective. Finally, given $f \in (B^1)^G$, define $\phi(Kg) = f(1^g, 1)$. Since $f \in (B^1)^G$ we have

$$\begin{aligned}0 = \partial f(1^{gh}, 1^h, 1) &= f(1^h, 1) - f(1^{gh}, 1) + f(1^{gh}, 1^h) \\ &= f(1^h, 1) - f(1^{gh}, 1) + f(1^g, 1) \\ &= \phi(Kh) - \phi(Kgh) + \phi(Kg)\end{aligned}$$

for all $g, h \in G$ so that $\phi \in Hom(G/K, F)$ satisfying $\partial\widehat{\phi} = f$ and θ is surjective. □

Now suppose $(L, *)$ is a loop of order $n = p^r m$ where $gcd(p, m) = 1$. Let $1 \in L$ be the identity, and let G be the left multiplication group of L; i.e. $G \leqslant Sym(L)$ is generated by the permutations $x \mapsto a * x$, $a \in L$. We show that the map $(f, g, h) \mapsto \partial f$ gives an isomorphism $\mathcal{V}_0(\mathcal{N}) \xrightarrow{\cong} (B^1)^G$. For $(f, g, h) \in \mathcal{V}_0(\mathcal{N})$ we have

$$f(x) + g(y) + h(x * y) = 0$$

for all $x, y \in L$ and so $f(x) = g(x) = -h(x)$ and

$$\begin{aligned}\partial f(a * x, a * y) &= f(a * x) - f(a * y) \\ &= f(a) + f(x) - f(a) - f(y) = f(x) - f(y) \\ &= \partial f(x, y)\end{aligned}$$

so that $\partial f \in (B^1)^G$. If $\partial f = 0$ then $f(x) = \partial f(x, 1) = 0$. Also if $\phi : L \to F$ such that $\partial\phi \in (B^1)^G$ then we easily check that $(f, f, -f) \in \mathcal{V}_0(\mathcal{N})$ where

$f(x) = \partial\phi(x,1) = \phi(x) - \phi(1)$:

$$\begin{aligned} f(x) + f(y) - f(x*y) &= \partial\phi(x,1) + \partial\phi(y,1) - \partial\phi(x*y,1) \\ &= \partial\phi(x,1) - \partial\phi(x*y,1) + \partial\phi(x*y,x) \\ &= \partial^2(x*y,x,1) = 0. \end{aligned}$$

Theorem 3.1 follows.

5. Third proof of main theorem (using finite fields)

We require the following well-known result, whose proof is included for completeness. As before we fix $F = \mathbb{F}_p$ where p is prime, and we use the convention that $0^0 = 1$.

Proposition 5.1. *Let $f : F \to F$, and for every $r \geqslant 0$, write $\sigma_{f,r} = \sum_{a\in F} f(a)^r \in F$. Then*

(*a*) *The map f is a permutation of F, if and only if*

$$\sigma_{f,0} = \sigma_{f,1} = \cdots = \sigma_{f,p-2} = 0 \quad \textit{and} \quad \sigma_{f,p-1} = -1.$$

(*b*) *We have $\sigma_{f,0} = \sigma_{f,1} = \cdots = \sigma_{f,p-2} = 0$, if and only if $\big|f(F)\big|$ equals either 1 or p.*

Proof. First suppose that the map f is a permutation of F, so that $\sigma_{f,r} = \sum_{a\in F} a^r$. Clearly $\sigma_{f,0} = p = 0 \in F$ and $\sigma_{f,p-1} = p{-}1 = -1 \in F$. Now suppose $1 \leqslant r \leqslant p{-}2$. For every $c \in \{1, 2, \ldots, p-1\}$ we have $c^r\sigma_{f,r} = \sum_{a\in F}(ca)^r = \sum_{a\in F} a^r = \sigma_{f,r}$ since the map $a \mapsto ca$ is a permutation of F. Now the polynomial $\sigma_{f,r}X^r - \sigma_{f,r} \in F[X]$ has $p{-}1 > r$ zeroes in the field F, so $\sigma_{f,r} = 0$ as required.

In the general case, for every $a \in F$, let $n_a = \big|f^{-1}(a)\big|$, so that $\sigma_{f,r} = \sum_{a\in F} a^r n_a$. The linear system

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 2 & \cdots & p{-}2 & p{-}1 \\ 0 & 1 & 2^2 & \cdots & (p{-}2)^2 & (p{-}1)^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 2^{p-2} & \cdots & (p{-}2)^{p-2} & (p{-}1)^{p-2} \\ 0 & 1 & 2^{p-1} & \cdots & (p{-}2)^{p-1} & (p{-}1)^{p-1} \end{bmatrix} \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ \vdots \\ n_{p-2} \\ n_{p-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix}$$

over $\mathbb{Q}$ has a unique solution, since the coefficient matrix is a nonsingular Vandermonde matrix. We have seen that $n_0 = n_1 = \cdots = n_{p-1} = 1$ is a solution, so (a) follows. Moreover the linear system

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 2 & \cdots & p-2 & p-1 \\ 0 & 1 & 2^2 & \cdots & (p-2)^2 & (p-1)^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 2^{p-2} & \cdots & (p-2)^{p-2} & (p-1)^{p-2} \end{bmatrix} \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ \vdots \\ n_{p-2} \\ n_{p-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

has as its general solution $n_0 = n_1 = \cdots = n_{p-1}$ since the coefficient matrix has rank $p-1$. Since $n_0 + n_1 + n_2 + \cdots + n_{p-1} = p$, we have either (i) $n_0 = n_1 = \cdots = n_{p-1} = 1$, or (ii) one of the n_k's is p and the others are zero. Conclusion (b) follows. □

Let $\mathcal{N}$ be a 3-net of odd prime order p, i.e. a set of p^2 triples $(x, y, z) \in F^3$ such that each point $(x, y, z) \in \mathcal{N}$ is uniquely determined by any two of its coordinates. We have $\operatorname{rank}_p \mathcal{N} = 3p-2-\dim \mathcal{V}_0$ where $\mathcal{V}_0$ is the set of all triples (f, g, h) of functions $F \to F$ such that $f(0) = g(0) = h(0) = 0$ and

$$f(x) + g(y) + h(z) = 0 \quad \text{for all } (x, y, z) \in \mathcal{N}.$$

We must show that $\dim \mathcal{V}_0 \leqslant 1$, and that equality holds iff the 3-net $\mathcal{N}$ is cyclic.

Suppose $(f, g, h) \in \mathcal{V}_0$ is nonzero. Using always the convention that $0^0 = 1$, we see that $\sigma_{f,0} = \sigma_{g,0} = \sigma_{h,0} = 0$. Note that for all $r \geqslant 0$ and all $(x, y, z) \in \mathcal{N}$, we have

$$h(z)^r = (-1)^r \sum_{s=0}^{r} \binom{r}{s} f(x)^{r-s} g(y)^s$$

by the Binomial Theorem. Summing over all p triples $(x, y, z) \in \mathcal{N}$ with a fixed value of y gives

$$\sigma_{h,r} = (-1)^r \sum_{s=0}^{r} \binom{r}{s} \sigma_{f,r-s} g(y)^s \tag{5.1}$$

for all $r \geqslant 0$, $y \in F$.

Summing (5.1) over all $y \in F$ gives

$$0 = \sum_{s=0}^{r} \binom{r}{s} \sigma_{f,r-s}\sigma_{g,s} \tag{5.2}$$

for all $r \geqslant 0$.

Theorem 5.2. *We have*

$$\sigma_{f,r} = \sigma_{g,r} = \sigma_{h,r} = \begin{cases} 0, & \text{for } r = 0,1,2,\ldots,p{-}2, \text{ and} \\ -1, & \text{for } r = p{-}1. \end{cases}$$

Proof. As previously noted, the result holds for $r = 0$. Assume the conclusion of the Theorem holds for all $r \in \{0,1,\ldots,t\}$ where $t \leqslant p{-}2$, and we will verify the conclusion in the case $r = t{+}1$. Applying (5.1) in the case $r = t{+}1$, we have $\sigma_{h,t+1} = (-1)^{t+1}\sigma_{f,t+1}$. Similarly, we obtain $\sigma_{f,t+1} = (-1)^{t+1}\sigma_{g,t+1}$ and $\sigma_{g,t+1} = (-1)^{t+1}\sigma_{h,t+1}$. Clearly the conclusion $\sigma_{f,t+1} = \sigma_{g,t+1} = \sigma_{h,t+1} = 0$ follows if t is even, but we proceed to obtain the same conclusion regardless of the parity of t.

We consider first the case $t \leqslant \frac{1}{2}(p{-}3)$. Applying (5.2) for $r = 2t{+}2$ yields

$$0 = \sum_{s=0}^{2t+2} \binom{2t{+}2}{s} \sigma_{f,2t+2-s}\sigma_{g,s} = \binom{2t{+}2}{t{+}1} \sigma_{f,t+1}\sigma_{g,t+1}\,.$$

Since $2t{+}2 < p$, this implies that $\sigma_{f,t+1}\sigma_{g,t+1} = 0$. Combining this with the previous paragraph yields $\sigma_{f,t+1} = \sigma_{g,t+1} = \sigma_{h,t+1} = 0$. Thus the conclusion holds for $r = t{+}1$ as well.

Next consider the case $\frac{1}{2}(p{-}1) \leqslant t < p{-}2$. Multiplying both sides of (5.1) by $g(y)^{2t+3-p}$ and setting $r = p{-}1$ yields

$$\begin{aligned} \sigma_{h,p-1}g(y)^{2t+3-p} &= \sum_{s=0}^{p-1} \binom{p{-}1}{s} \sigma_{f,p-1-s}g(y)^{s+2t+3-p} \\ &= \sum_{s=0}^{p-t-2} \binom{p{-}1}{s} \sigma_{f,p-1-s}g(y)^{s+2t+3-p}. \end{aligned}$$

Note that $2t{+}3{-}p \geqslant 2$, so all exponents are non-negative. Now observe

that $2t+3-p < t$ and sum over $y \in F$ to obtain

$$0 = \sigma_{h,p-1}\sigma_{g,2t+3-p} = \sum_{s=0}^{p-t-2} \binom{p-1}{s} \sigma_{f,p-1-s}\sigma_{g,s+2t+3-p}$$
$$= \binom{p-1}{p-t-2} \sigma_{f,t+1}\sigma_{g,t+1}.$$

Since the latter binomial coefficient is not divisible by p, we obtain $\sigma_{f,t+1}\sigma_{g,t+1} = 0$. This yields $\sigma_{f,t+1} = \sigma_{g,t+1} = \sigma_{h,t+1} = 0$ as before.

Applying (5.1) for $r = p-1$ gives $\sigma_{h,p-1} = \sigma_{f,p-1}$; and similarly, $\sigma_{h,p-1} = \sigma_{g,p-1}$. By assumption, $(f,g,h) \in \mathcal{V}_0$ is nonzero; therefore by Proposition 5.1 we have $\sigma_{f,p-1} = \sigma_{g,p-1} = \sigma_{h,p-1} = -1$ and each of the maps f, g, h is a permutation of F. We may assume that $f(x) = g(x) = -h(x) = x$ for all $x \in F$; otherwise relabel the lines in each parallel class so that this is the case. Since $f(x) + g(y) + h(z) = 0$ for all $(x,y,z) \in \mathcal{N}$, we obtain $\mathcal{N} = \{(x, y, x+y) : x, y \in F\}$ and so the 3-net $\mathcal{N}$ is cyclic. □

6. Fourth proof of main theorem (using exponential sums)

Let $F = \mathbb{F}_p$ where p is prime, and let $\zeta \in \mathbb{C}$ be a primitive p-th root of unity. We have a well-defined map

$$e : F \to \mathbb{Z}[\zeta], \quad a \mapsto \zeta^a$$

satisfying $e(a+b) = e(a)e(b)$ for all $a, b \in F$. Each function $f : F \to F$ gives rise to an *exponential sum*

$$S_f = \sum_{a \in F} e(f(a)) \in \mathbb{Z}[\zeta].$$

Now suppose $\mathcal{N}$ is a 3-net of order p, and let $(f,g,h) \in \mathcal{V}_0(\mathcal{N})$. Summing $\zeta^{f(a)+g(b)} = \zeta^{-h(c)}$ over all $(a,b,c) \in \mathcal{N}$ gives $S_f S_g = \overline{S_h}$, and similarly $S_g S_h = \overline{S_f}$ and $S_h S_f = \overline{S_g}$. Thus

$$|S_f|^2 = |S_g|^2 = |S_h|^2 = \tfrac{1}{p} S_f S_g S_h.$$

Now if $|S_f| = |S_g| = |S_h| = p$ then $f, g, h : F \to F$ are constant functions, but then the condition $f(0) = g(0) = h(0) = 0$ forces $(f,g,h) = (0,0,0)$.

Otherwise we must have $S_f = S_g = S_h = 0$, so that $f, g, h : F \to F$ are permutations. After permuting labels, we may assume that

$$f(X) = X, \quad g(X) = X, \quad h(X) = -X.$$

Now

$$0 = f(a) + g(b) + h(c) = a + b - c$$

for all $(a, b, c) \in \mathcal{N}$, i.e.

$$\mathcal{N} = \{(a, b, a+b) : a, b \in F\}$$

which is the cyclic 3-net of order p. □

7. 4-nets of prime order

Let $\mathcal{N}$ be a 4-net of prime order p, and let $(f, g, h, u) \in \mathcal{V}(\mathcal{N})$. In the notation of Section 6, we sum the quantity $\zeta^{f(x)+g(y)} = \zeta^{-h(z)-u(t)}$ over all $(x, y, z, t) \in \mathcal{N}$ to obtain $S_f S_g = \overline{S_h S_u}$. It is not hard to check that either

$$|S_f| = |S_g| = |S_h| = |S_u| > 0$$

or at least three of the exponential sums $\{S_f, S_g, S_h, S_u\}$ vanish, in which case the corresponding members of $\{f, g, h, u\}$ are permutations. With some further investigation we have shown

Theorem 7.1. [8] *Let $\mathcal{N}$ be a 4-net of order p. Then:*

(*i*) *The number of cyclic 3-subnets of $\mathcal{N}$ is 0, 1, 3 or 4.*

(*ii*) *$\mathcal{N}$ has four cyclic 3-subnets iff $\mathcal{N}$ is Desarguesian.*

(*iii*) *Suppose $\mathcal{N}$ has at least one cyclic 3-subnet. Then $\mathcal{N}$ has rank at least $4p-3$, and equality holds iff $\mathcal{N}$ is Desarguesian.*

We remark that (*i*) and (*ii*) are best possible in the sense that there exist (necessarily non-Desarguesian) 4-nets of prime order p having 0, 1 or 3 cyclic subnets. Examples of these for $p = 7, 11$ are found at [6,7].

References

[1] **V. D. Belousov**: MR561712 (81g:20133), 1981. Review of *Kvasigrupe (Quasigroups)* by U. Janez, 1979.

[2] **R. H. Bruck and H. J. Ryser**: *The nonexistence of certain finite projective planes*, Canad. J. Math. **1** (1949), 88 – 93.

[3] **S. S. Chern and P. Griffiths**: *Abel's theorem and webs*, Jahresberichte der Deut. Math. Ver. **80** (1978), 13 – 110; also, *Corrections and addenda to our paper: Abel's theorem and webs*, same Journal, **83** (1981), 78 – 83.

[4] **P. A. Griffiths**: *Variations on a theorem of Abel*, Inventiones Math. **35** (1976), 321 – 390.

[5] **G. E. Moorhouse**: *Bruck nets, codes, and characters of loops*, Des. Codes Cryptogr. **1** (1991), 7 – 29.

[6] **G. E. Moorhouse**: *Nets and Latin squares of order* 7, available at `http://www.uwyo.edu/moorhouse/pub/nets7/`.

[7] **G. E. Moorhouse**: *Nets and Latin squares of order* 11, available at `http://www.uwyo.edu/moorhouse/pub/nets11/`.

[8] **G. E. Moorhouse**: *Ranks of Nets and of Webs*, preliminary version, 2005.

Department of Mathematics
University of Wyoming
1000 E. University Ave.
Laramie, WY 82071
U.S.A.
e-mail: moorhous@uwyo.edu

Received January 2, 2006

Quasigroups and Related Systems **14** (2006), 73 – 80

A short basis for the variety of WIP PACC-loops

J. D. Phillips

Abstract

The variety of weak inverse property, power-associative, conjugacy closed loops (WIP PACC-loops) is an ideal variety from which to investigate CC-loops. We give a surprisingly short basis for this variety. We also give a useful associator identity.

1. Introduction

A *quasigroup* $(Q, \cdot)$ is a set Q with a binary operation $\cdot$ such that for each $a, b \in Q$ the equations $a \cdot x = b$ and $y \cdot a = b$ have unique solutions $x, y \in Q$. A *loop* is a quasigroup with a two-sided neutral element, 1. We write xy instead of $x \cdot y$, and stipulate that $\cdot$ have lower priority than juxtaposition among factors to be multiplied—for instance, $x \cdot yz$ stands for $x(yz)$. For an overview of the theory of loops, see [3, 4, 17].

A loop is *conjugacy closed* (a *CC-loop*) iff it satisfies both of the following equations:

$$xy \cdot z = xz \cdot (z \backslash (yz)) \quad \text{(RCC)} \qquad\qquad z \cdot yx = ((zy)/z) \cdot zx \quad \text{(LCC)}$$

This definition is owing to Goodaire and Robinson [10, 11]; CC-loops were introduced independently, with different terminology, by Сойкис [20]. Further development of the theory can be found in [1, 2, 5, 6, 7, 14, 15]. The literature is not uniform as to which of these two equations is left (LCC) and which is right (RCC). With our choice here LCC is equivalent to requiring that the set of *left* multiplication maps be closed under conjugation. In [14, 15], the equation labels LCC and RCC were arranged in the opposite order.

The CC-loops which are also *diassociative* (that is every $\langle x, y \rangle$ is a group) are the *extra loops* introduced by Fenyves [8, 9]. For these, a detailed structure theory was described in [12], and their relationship to the other loops (and quasigroups) of so-called Bol-Moufang type is detailed in [18, 19].

2000 Mathematics Subject Classification: 20N05
Keywords: conjugacy closed loop, power-associative, weak inverse property

Definition 1.1. For any loop Q and for $c \in Q$:

1. Define c^ρ and c^λ by: $cc^\rho = c^\lambda c = 1$.
2. c is *power-associative* iff the subloop $\langle c \rangle$ is a group. Q is *power-associative* iff every element is power-associative.

The two parts of this definition are related by:

Lemma 1.2. (cf. [15]) *Let c be an element of a CC-loop Q. Then*

$$c \text{ is power associative} \quad \text{iff} \quad c^\rho = c^\lambda \quad \text{iff} \quad cc^2 = c^2c$$

Power-associative CC-loops (PACC-loops) were thoroughly analyzed by Kinyon and Kunen in [13]. Central to their analysis is the notion of weak inverse property elements.

Definition 1.3. An element c of a loop Q is a *weak inverse property (WIP) element* iff $\forall x \in Q$,

$$c(xc)^\rho = x^\rho \qquad (cx)^\lambda c = x^\lambda \,. \tag{WIP}$$

A loop Q is then a *weak inverse property loop* if all of its elements have the weak inverse property.

Kinyon and Kunen showed that if Q is a PACC-loop and if $c \in Q$, then $c^{12} \in N(Q)$, the nucleus of Q; they prove this by showing that c^3 is a WIP element and c^6 is an extra element.

There is thus, a chain of five prominent varieties of CC-loops: (1) groups, (2) extra loops, (3) WIP PACC-loops, (4) PACC-loops, (5) CC-loops.

Obviously, a great deal has been written about the varieties (1), (2), and (5). And with the appearance of [13], the variety in (4) has now been analyzed in great detail; moreover, the prominence of the weak inverse property in the theory of CC-loops is now apparent. The variety in (3) is thus, a natural setting in which to investigate CC-loops: rich enough for deep structural theorems, but still, quite general. Unfortunately, the equations that axiomatize this variety are unwieldy. The main purpose of this paper is to give a surprisingly short and elegant basis for the variety of WIP PACC-loops.

2. Main result

Theorem 2.1. *In the variety of loops, WIP PACC-loops are axiomatized by the identities*

$$(xy \cdot x) \cdot xz = x \cdot ((yx \cdot x)z) \tag{LWPC}$$

and

$$zx \cdot (x \cdot yx) = (z(x \cdot xy)) \cdot x. \tag{RWPC}$$

Proof. We first show that WIP PACC-loops satisfy both LWPC and RWPC. We proceed in four steps. Firstly, we note that in WIP PACC loops, it is easy to see that both $(x/y)\cdot yz = y\cdot((y\backslash x)z)$ and $(x/y)\backslash 1 = y(x\backslash 1)$ hold, and that together they imply $x \cdot ((x\backslash y)z) = ((y\backslash 1)\backslash(x\backslash 1)) \cdot ac$. Secondly, we note that in WIP PACC loops, it is easy to see that $(x\backslash y)\ 1 = (y\backslash 1)x$, $((x\backslash y)\backslash 1)x \cdot (x\backslash y) = (x\backslash y)\backslash y$, and $x/y = (x\backslash 1)\backslash(y\backslash 1)$ hold, and that together they imply $(x\backslash y)\backslash y = (y\backslash 1) \cdot xy$. Thirdly, we note that in WIP PACC loops, it is easy to see that $(xy)\backslash 1 = y\backslash(x\backslash 1)$, $(x\backslash y)\backslash y = (y\backslash 1) \cdot xy$, and $(x\backslash y)z = ((y\backslash 1)\backslash(x\backslash 1) \cdot xz)$ hold, and that together they imply $(x\backslash 1)(xy\cdot x) = yx$. Fourthly, and finally, we note that in WIP PACC loops, it is easy to see that both $xy\backslash 1 = y\backslash(x\backslash 1)$ and $x\backslash(yx) = (x\backslash 1)y) \cdot x$ hold and that together the three identities just established, implies LWPC. Of course, dualize this proof to obtain RWPC.

We now show that loops satisfying both LWPC and RWPC are, in fact, WIP PACC loops. In LWPC, replace z with $x\backslash z$ to obtain $(xy\cdot x)z = x((yx\cdot x)\cdot(x\backslash z))$. Now, replace y with $(y/x)/x$ to obtain $(x((y/x)/x)\cdot x)z = x(y \cdot (x\backslash z))$. In other words, there exists a function $f(x, y)$ such that $L(x)^{-1}L(y)L(x) = L(f(x, y))$. Thus, in the variety of loops, LWPC implies LCC. Similarly, RWPC implies RCC. Also, it is easy to see that LWPC implies $x^\rho = x^\lambda$, which by Lemma 1.2 guarantees power-associativity. Finally, we offer an Otter output file proving that LWPC and RWPC together imply WIP:

```
Length of proof is 89.  Level of proof is 20.
---------------- PROOF ----------------
240 [] x=x.
242,241 [] 1*x=x.
244,243 [] x*1=x.
245 [] ((x*y)*x)*(x*z)=x*(((y*x)*x)*z).
246 [] (x*y)*(y*(z*y))=(x*(y*(y*z)))*y.
247 [copy,246,flip.1] (x*(y*(y*z)))*y=(x*y)*(y*(z*y)).
250,249 [] 1/x=x\1.
252,251 [] x*(x\y)=y.
254,253 [] x\(x*y)=y.
255 [] (x*y)/y=x.
```

```
258,257 [] (x/y)*y=x.
259 [] ((x*y)/x)*(x*z)=x*(y*z).
261 [] (x*y)*(y\(z*y))=(x*z)*y.
262 [] A*((B*A)\1)!=B\1.
263 [copy,245,flip.1] x*(((y*x)*x)*z)=((x*y)*x)*(x*z).
264 [copy,261,flip.1] (x*y)*z=(x*z)*(z\(y*z)).
265 [para_into,245.1.1.1.1,243.1.1,demod,242]
    (x*x)*(x*y)=x*((x*x)*y).
267 [para_into,245.1.1.2,243.1.1,demod,244]
    ((x*y)*x)*x=x*((y*x)*x).
271 [para_into,247.1.1.1,241.1.1,demod,242]
    (x*(x*y))*x=x*(x*(y*x)).
274,273 [para_into,251.1.1,241.1.1] 1\x=x.
285 [para_into,253.1.1.2,243.1.1] x\x=1.
287 [para_into,255.1.1.1,251.1.1] x/(y\x)=y.
298,297 [para_into,257.1.1.1,249.1.1] (x\1)*x=1.
302,301 [para_from,257.1.1,253.1.1.2] (x/y)\x=y.
305 [para_into,259.1.1.1.1,251.1.1] (x/y)*(y*z)=y*((y\x)*z).
309 [para_into,259.1.1.2,251.1.1] ((x*y)/x)*z=x*(y*(x\z)).
310 [copy,309,flip.1] x*(y*(x\z))=((x*y)/x)*z.
315,314 [para_into,261.1.1.2.2,257.1.1,flip.1]
    (x*(y/z))*z=(x*z)*(z\y).
316 [para_into,261.1.1.2.2,251.1.1]
    (x*(y\z))*((y\z)\z)=(x*y)*(y\z).
326 [para_into,263.1.1.2,251.1.1,flip.1]
    ((x*y)*x)*(x*(((y*x)*x)\z))=x*z.
343 [para_into,265.1.1.2,251.1.1,flip.1] x*((x*x)*(x\y))=(x*x)*y.
346,345 [para_into,265.1.1.2,243.1.1,demod,244] (x*x)*x=x*(x*x).
359 [para_from,267.1.1,255.1.1.1] (x*((y*x)*x))/x=(x*y)*x.
369 [para_into,271.1.1.1.2,251.1.1,flip.1]
    x*(x*((x\y)*x))=(x*y)*x.
409,408 [para_into,287.1.1,249.1.1] (x\1)\1=x.
423,422 [para_from,297.1.1,261.1.1.2.2,flip.1]
    (x*(y\1))*y=(x*y)*(y\1).
425 [para_from,297.1.1,264.1.1.1,demod,242,flip.1]
    ((x\1)*y)*(y\(x*y))=y.
428,427 [para_from,297.1.1,261.1.1.1,demod,242]
    x\(y*x)=((x\1)*y)*x.
430 [back_demod,425,demod,428] ((x\1)*y)*(((y\1)*x)*y)=y.
453 [para_from,314.1.1,267.1.1.1,demod,258]
    ((x*x)*(x\y))*x=x*(y*x).
512,511 [para_into,359.1.1.1.2.1,257.1.1,demod,315]
    (x*(y*x))/x=(x*x)*(x\y).
554,553 [para_into,422.1.1.1,297.1.1,demod,242,409,flip.1]
```

```
    (x*x)*(x\1)=x.
560,559 [para_from,427.1.1,251.1.1.2] x*(((x\1)*y)*x)=y*x.
565 [para_from,430.1.1,255.1.1.1] x/(((x\1)*y)*x)=(y\1)*x.
572,571 [para_into,453.1.1.1.2,253.1.1] ((x*x)*y)*x=x*((x*y)*x).
574,573 [para_into,453.1.1,422.1.1,demod,346,242]
    (x*(x*x))*(x\1)=x*x.
579 [para_from,553.1.1,427.1.1.2,demod,409,574] (x\1)\x=x*x.
599 [para_into,559.1.1.2.1.1,408.1.1] (x\1)*((x*y)*(x\1))=y*(x\1).
602,601 [para_into,559.1.1.2.1,559.1.1,demod,423,409,423,flip.1]
    ((x*y)*x)*(x\1)=x*((y*x)*(x\1)).
603 [para_into,559.1.1.2.1,251.1.1,flip.1] ((x\1)\y)*x=x*(y*x).
624,623 [para_into,579.1.1.1,408.1.1] x\(x\1)=(x\1)*(x\1).
636,635 [para_from,603.1.1,255.1.1.1,demod,512,flip.1]
    (x\1)\y=(x*x)*(x\y).
646,645 [para_into,305.1.1.1,249.1.1] (x\1)*(x*y)=x*((x\1)*y).
678 [para_into,635.1.1,253.1.1,flip.1] (x*x)*(x\((x\1)*y))=y.
680 [para_from,635.1.1,287.1.1.2] x/((y*y)*(y\x))=y\1.
684,683 [para_from,635.1.1,251.1.1.2] (x\1)*((x*x)*(x\y))=y.
686 [para_into,645.1.1.2,343.1.1,demod,684] (x\1)*((x*x)*y)=x*y.
692,691 [para_into,645.1.1.2,251.1.1,flip.1]
    x*((x\1)*(x\y))=(x\1)*y.
702,701 [para_from,645.1.1,253.1.1.2,demod,636,254]
    (x*x)*((x\1)*y)=x*y.
710,709 [para_into,309.1.1,243.1.1] (x*y)/x=x*(y*(x\1)).
720 [back_demod,310,demod,710] x*(y*(x\z))=(x*(y*(x\1)))*z.
725,724 [para_into,678.1.1.2.2.1,635.1.1,demod,554,636,254]
    ((x\1)*(x\1))*((x*x)*y)=y.
728 [para_from,678.1.1,645.1.1.2,flip.1]
    (x*x)*(((x*x)\1)*(x\((x\1)*y)))=((x*x)\1)*y.
733,732 [para_from,678.1.1,253.1.1.2] (x*x)\y=x\((x\1)*y).
734 [back_demod,728,demod,733,244,624,733,244,624]
    (x*x)*(((x\1)*(x\1))*(x\((x\1)*y)))=((x\1)*(x\1))*y.
740 [para_into,680.1.1.2.2,301.1.1,demod,315,258]
    x/(x*(y\x))=(x/y)\1.
754 [para_from,686.1.1,265.1.1.2,demod,725]
    ((x\1)*(x\1))*(x*y)=(x\1)*y.
756 [para_into,316.1.1.1,251.1.1] x*((y\x)\x)=(y*y)*(y\x).
766,765 [para_into,701.1.1.2,343.1.1,demod,636,725,252]
    (x*x)*(((x\1)*(x\1))*y)=y.
769,768 [back_demod,734,demod,766] x\((x\1)*y)=((x\1)*(x\1))*y.
771,770 [back_demod,732,demod,769] (x*x)\y=((x\1)*(x\1))*y.
784 [para_into,720.1.1.2.2,285.1.1,demod,244,flip.1]
    (x*(y*(x\1)))*x=x*y.
816 [para_into,740.1.1.2.2,301.1.1,flip.1] (x/(x/y))\1=x/(x*y).
```

```
828,827 [para_from,754.1.1,247.1.1.1,demod,646,423,298,242,646]
    (x*((x\1)*y))*x=x*((x\1)*(y*x)).
843,842 [para_from,756.1.1,253.1.1.2,flip.1]
    (x\y)\y=y\((x*x)*(x\y)).
858,857 [para_from,784.1.1,427.1.1.2,demod,254,646,828,423,flip.1]
    x*((x\1)*((y*x)*(x\1)))=y.
862 [para_from,784.1.1,253.1.1.2] (x*(y*(x\1)))\(x*y)=x.
867 [para_from,816.1.1,287.1.1.2,demod,250,flip.1]
    x/(x/y)=(x/(x*y))\1.
889 [para_into,862.1.1.2,251.1.1] (x*((x\y)*(x\1)))\y=x.
894,893 [para_from,867.1.1,301.1.1.1,demod,636,302,315,258,flip.1]
    x/y=x*((x*y)\x).
911 [back_demod,565,demod,894,560] x*((y*x)\x)=(y\1)*x.
925 [back_demod,257,demod,894] (x*((x*y)\x))*y=x.
935 [para_into,889.1.1.1.2.1,635.1.1,demod,843,554,274,572,252,646]
    (x*((x\1)*(y*x)))\y=x\1.
951 [para_into,925.1.1.1.2.1,251.1.1] (x*(y\x))*(x\y)=x.
959 [para_into,935.1.1.1.2.1,889.1.1,demod,692,244,692,244,646,692,
    244,843,554,274] (x*((x\1)*(y*(x\1))))\y=x.
995 [para_from,951.1.1,720.1.1.2,flip.1]
    (x*((x*(y\x))*(x\1)))*y=x*x.
1011 [para_from,369.1.1,599.1.1.2.1,demod,602,646,858,flip.1]
     (x*((x\y)*x))*(x\1)=y.
1025 [para_from,911.1.1,253.1.1.2,demod,428,flip.1]
     (x*y)\y=((y\1)*(x\1))*y.
1034,1033 [para_into,1011.1.1.1.2.1,253.1.1] (x*(y*x))*(x\1)=x*y.
1048 [para_into,1033.1.1.1,559.1.1] (x*y)*(y\1)=y*((y\1)*x).
1050 [para_into,1033.1.1.2,889.1.1,demod,692,244,692,244,692,244]
     (x\1)*(y*(x\1)))*x=(x\1)*y.
1068 [para_into,1048.1.1.2,770.1.1,demod,244,771,244,766]
     (x*(y*y))*((y\1)*(y\1))=x.
1080 [para_from,1050.1.1,326.1.1.2.2.1.1,demod,828,423,858]
     x*(y*((((y\1)*x)*y)\z))=y*z.
1092 [para_into,1080.1.1.2.2,1025.1.1,demod,560]
     x*((((y\1)*x)\1)*y)=y*y.
1098 [para_into,1092.1.1.2.1.1,251.1.1,demod,636]
     ((x*x)*(x\y))*((y\1)*x)=x*x.
1102 [para_into,1098.1.1.2.1,889.1.1,demod,692,244]
     ((x*x)*(x\(y\1)))*(y*x)=x*x.
1110 [para_from,1102.1.1,935.1.1.1.2.2]
     ((x*y)*(((x*y)\1)*(y*y)))\((y*y)*(y\(x\1)))=(x*y)\1.
1257 [para_into,1068.1.1.2.1,959.1.1,demod,242,692,244,242,692,244,
     242,692,244,843,554,274] (x*((y\1)*(y\1)))*(y*y)=x.
1263 [para_into,1257.1.1.1,995.1.1,demod,771,843,554,274,843,554,
```

```
     274,1034] (x*x)*(y*y)=x*(x*(y*y)).
1266,1265 [para_from,1263.1.1,765.1.1.2,demod,702]
     x*((x\1)*(y*y))=y*y.
1272,1271 [back_demod,1110,demod,1266,254,flip.1] (x*y)\1=y\(x\1).
1295 [back_demod,262,demod,1272,252] B\1!=B\1.
1296 [binary,1295.1,240.1] $F.
------------ end of proof -------------
```

Problem 2.2. *Is a quasigroup that satisfies* LWPC *and* RWPC *a loop?*

Recall, that an element c of a CC-loop is a *Moufang element* iff for all x, y $(c \cdot xy)c = cx \cdot yc$ (see [4],VII§2), and c is an *extra element* iff for all x, y $c(x \cdot yc) = (cx \cdot y)c$ [13]. Also, recall that the *associator* (x, y, z) is given by $xy \cdot z = (x \cdot yz) \cdot (x, y, z)$. Finally, given a loop L and an element $x \in L$, define the right and left multiplications by $xy = xR_y = yL_x$. We then may define the *inner mappings* on L by:

$$R(x, y) := R_x R_y R_{xy}^{-1}, \qquad L(x, y) := L_x L_y L_{yx}^{-1}, \qquad T_x := R_x L_x^{-1}.$$

Theorem 2.3. *Let c be a Moufang element in a CC-loop L. Then $\forall x, y$ we have $(c, x, y)^2 = (c^2, x, y)$.*

Proof. The otter output file of this proof may be found here:
`http://persweb.wabash.edu/facstaff/phillipj/research.html` □

Let c be an extra element in a CC-loop. Then $\forall x, y$ we have $(c, x, y)^2 = 1$ [13]. We now give a version of this result for Moufang elements.

Corollary 2.4. *Let c be a Moufang element in a WIP CC-loop. Then $\forall x, y$ we have $(c, x, y)^2 = 1$.*

Proof. In WIP CC loops, squares are nuclear [14]. Now, use Theorem 2.3. □

Acknowledgement. Our investigations were aided by the automated reasoning tool OTTER [16]. We thank Michael Kinyon for his helpful and generous comments.

References

[1] **А. С. Басараб**: *Об одном классе G-луп*, Математические Исследования Том **3**, Вып. 2 (8) (1968), 72 – 77.

[2] **А. С. Басараб**: *Класс* LK-*луп*, Математические Исследования, Вып. **120** (1991), 3 – 7.

[3] **В. Д. Белоусов**: Основы Теории Квазигрупп и Луп, Издательство «Наука», Москва, 1967.

[4] **R. H. Bruck**: A Survey of Binary Systems, Springer, 1971.

[5] **P. Csörgő and A. Drápal**: *Left conjugacy closed loops of nilpotency class* 2, preprint.

[6] **A. Drápal**: *Conjugacy closed loops and their multiplication groups*, J. Algebra **272** (2004), 838 – 850.

[7] **A. Drápal**: *Structural interactions of conjugacy closed loops*, preprint.

[8] **F. Fenyves**: *Extra loops I*, Publ. Math. Debrecen **15** (1968), 235 – 238.

[9] **F. Fenyves**: *Extra loops II*, Publ. Math. Debrecen **16** (1969), 187 – 192.

[10] **E. G. Goodaire and D. A. Robinson**: *A class of loops which are isomorphic to all loop isotopes*, Canadian J. Math. **34** (1982), 662 – 672.

[11] **E. G. Goodaire and D. A. Robinson**: *Some special conjugacy closed loops*, Canadian Math. Bull. **33** (1990) 73 – 78.

[12] **M. K. Kinyon and K. Kunen**: *The structure of extra loops*, Quasigroups and Related Systems **12** (2004), 39 – 60.

[13] **M. K. Kinyon and K. Kunen**: *Power-associative conjugacy closed loops*, J. Algebra, to appear.

[14] **M. K. Kinyon, K. Kunen, and J. D. Phillips**: *Diassociativity in conjugacy closed loops*, Comm. Algebra **32** (2004), 767 – 786.

[15] **K. Kunen**: *The structure of conjugacy closed loops*, Trans. Amer. Math. Soc. **352** (2000), 2889 – 2911.

[16] **W. W. McCune**: *OTTER 3.3 Reference Manual and Guide*, Argonne National Laboratory Technical Memorandum ANL/MCS-TM-263, 2003; `http://www.mcs.anl.gov/AR/otter/`

[17] **H. O. Pflugfelder**: Quasigroups and Loops: Introduction, Sigma Series in Pure Math. **8**, Heldermann, 1990.

[18] **J.D. Phillips and P. Vojtěchovský**: *The varieties of loops of Bol-Moufang type*, Algebra Universalis, to appear.

[19] **J.D. Phillips and P. Vojtěchovský**: *The varieties of quasigroups of Bol-Moufang type*, J. Algebra **293** (2005), 17 – 33.

[20] **Л. Р. Сойкис**: *О специальных лупах*, in Вопросы Теории Квазигрупп и Луп (В. Д. Белоусов, ed.), Редакц.-Издат. Отдел Акад. Наук Молдав. ССР, Кишинев, 1970, 122 – 131.

Received March 20, 2006

Department of Mathematics & Computer Science, Wabash College, Crawfordsville, IN 47933, U.S.A.
e-mail: phillipj@wabash.edu

Quasigroups and Related Systems **14** (2006), 81 – 89

Duality for central piques

Anna B. Romanowska and Jonathan D. H. Smith

Abstract

A duality for locally compact central piques is established, based on Pontryagin duality for locally compact abelian groups. The duality restricts to yield Suvorov duality for quasigroup modes.

1. Introduction

Piques are quasigroups with a pointed idempotent element. As such, they form an intermediate class between loops and general quasigroups. A quasigroup Q is *central* if the diagonal $\widehat{Q}$ is a normal subquasigroup of the square Q^2, or an equivalence class of a quasigroup congruence relation on Q^2 [1, Ch. III]. Each central quasigroup Q is very tightly related to a central pique, namely the quotient $Q^2/\widehat{Q}$ with $\widehat{Q}$ as the pointed idempotent. (The exact relationship is central isotopy [1, Ch. III]. In particular, $Q \times Q$ and $Q \times Q^2/\widehat{Q}$ are isomorphic.) Moreover, the class of central piques includes many of the quasigroups encountered in practice, such as abelian groups under subtraction or the standard constructions of orthogonal quasigroups based on finite fields. It also includes the class of *quasigroup modes*, idempotent and entropic quasigroups [3].

As part of a general determination of the character tables of finite (discrete) central piques, the paper [4] sketched a duality for them, and raised the problem of

2000 Mathematics Subject Classification: 20N05
Keywords: quasigroup, loop, centrality, duality, topological quasigroup, Pontryagin duality
The second author gratefully acknowledges the hospitality of the Faculty of Mathematics and Information Sciences of Warsaw University of Technology during the completion of this paper while he was on a Faculty Professional Development Assignment from Iowa State University.

extension of the duality theory from finite (discrete) central piques to general locally compact central piques.

The purpose of the current paper is to present such an extension (Theorem 5.5). Earlier, Suvorov exhibited a duality for locally compact topological quasigroup modes on the basis of Pontryagin duality [7]. Suvorov duality may now be seen as a special case of the duality for central piques (Section 6).

Readers are referred to [6] for concepts and notational conventions not otherwise defined explicitly in this paper.

2. Quasigroups and piques

A *quasigroup* $(Q, \cdot)$ is considered combinatorially as a set Q equipped with a binary *multiplication* operation denoted by $\cdot$ or simple juxtaposition of the two arguments, in which specification of any two of x, y, z in the equation $x \cdot y = z$ determines the third uniquely. Equationally, a quasigroup $(Q, \cdot, /, \backslash)$ is a set Q equipped with three binary operations of multiplication, *right division* $/$ and *left division* $\backslash$, satisfying the identities:

$$\begin{array}{ll} \text{(IL)} & y \setminus (y \cdot x) = x; \\ \text{(IR)} & x = (x \cdot y)/y; \\ \text{(SL)} & y \cdot (y \setminus x) = x; \\ \text{(SR)} & x = (x/y) \cdot y. \end{array}$$

(Suppressing explicit mention of the division operations of a quasigroup, one may denote it merely as $(Q, \cdot)$ instead.) The equational definition of quasigroups means that they form a variety in the sense of universal algebra, and are thus susceptible to study by the concepts and methods of universal algebra [6]. In particular, a quasigroup $(Q, \cdot, /, \backslash)$ is *topological* if the underlying set Q is a topological space, and if the operations $Q^2 \to Q; (x, y) \mapsto x \cdot y$, $Q^2 \to Q; (x, y) \mapsto x/y$ and $Q^2 \to Q; (x, y) \mapsto x \backslash y$ are continuous maps from the product space Q^2.

An element e of a quasigroup Q is said to be *idempotent* if $\{e\}$ forms a singleton subquasigroup of Q. A *pique* or *pointed idempotent quasigroup* [1, §III.5] is a quasigroup P, containing an idempotent element 0, that has its quasigroup structure of multiplication and the divisions enriched by a nullary operation selecting the idempotent element 0. Note that piques also form a variety.

For each element q of a quasigroup $(Q, *)$, the *right multiplication* $R_*(q)$ or

$$R(q) : Q \to Q; x \mapsto x * q$$

and *left multiplication* $L_*(q)$ or

$$L(q) : Q \to Q; x \mapsto q * x$$

are elements of the group $Q!$ of bijections from the set Q to itself. The subgroup of $Q!$ generated by all the right and left multiplications is called the *multiplication group* $\operatorname{Mlt} Q$ of Q. For a pique P with pointed idempotent 0, it is conventional to set $R = R(0)$ and $L = L(0)$. The stabilizer of 0 in the permutation group $\operatorname{Mlt} P$ is called the *inner multiplication group* $\operatorname{Inn} P$ of P. For example, if P is a group, then the inner multiplication group of the pique P is just the inner automorphism group of the group P.

3. Central piques

A *loop* L is a pique in which the pointed idempotent element 1 acts as an identity. For a general pique $(P, \cdot, 0)$, the *cloop* or *corresponding loop* is the loop $B(P)$ or $(P, +, 0)$ in which the "multiplication" operation $+$ is defined by

$$x + y = xR^{-1} \cdot yL^{-1}. \tag{3.1}$$

Inverting (3.1), the multiplication of a pique is recovered from the cloop by

$$x \cdot y = xR + yL. \tag{3.2}$$

Definition 3.1. A pique $(P, \cdot, 0)$ is said to be *central*, or to lie in the class $\mathfrak{Z}$, if $B(P)$ is an abelian group, and $\operatorname{Inn} P$ is a group of automorphisms of $B(P)$.

Remark 3.2. The syntactical Definition 3.1 of pique centrality is chosen for its concreteness, and because it is well suited to the purposes of the current paper. The equivalence of this definition with the structural characterization given in the introduction is discussed in [1, §III.5], [5, §3.5].

In a central pique P, the right division is given by

$$x/y = (x - yL)R^{-1} \tag{3.3}$$

and the left division by

$$x\backslash y = (y - xR)L^{-1} \tag{3.4}$$

for x, y in P, using the subtraction of the cloop $B(P)$. Topological central piques are easily characterized.

Proposition 3.3. *A central pique P is topological if and only if the cloop $B(P)$ is a topological abelian group, and the maps $R : P \to P$ and $L : P \to P$ are homeomorphisms.*

Proof. If P is topological, then the maps $R : P \to P; x \mapsto x \cdot 0$ and $L : P \to P; x \mapsto 0 \cdot x$ are certainly continuous, as are their respective inverses $R^{-1} : P \to P; x \mapsto x/0$ and $L^{-1} : P \to P; x \mapsto 0\backslash x$. The cloop addition (3.1) is then continuous, as is the negation $-y = 0/(yRL^{-1})$ according to (3.3).

Conversely, if the cloop is a topological group and the maps R, L are homeomorphisms, then it is apparent from (3.1), (3.3) and (3.4) that the pique is topological. □

Let P be a central pique. Let $\langle R, L\rangle$ denote the free group on the 2-element set $\{R, L\}$. Then the group homomorphism

$$\langle R, L\rangle \to \operatorname{Inn} P; \quad R \mapsto R(0), \quad L \mapsto L(0)$$

makes P a right $\langle R, L\rangle$-module. Conversely, a right $\langle R, L\rangle$-module P yields a pique with multiplication given by (3.2).

Proposition 3.4. *Let $\langle R, L\rangle$ denote the free group on the 2-element set $\{R, L\}$.*

(a) *The category of central piques is equivalent to the category of right $\langle R, L\rangle$-modules.*

(b) *The category of locally compact central piques is equivalent to the category of locally compact right $\langle R, L\rangle$-modules.*

Proof. Statement (a) follows immediately from the preceding considerations, while (b) holds by Proposition 3.3. □

4. Schizophrenic objects and diagrammatic duality

Let $\mathbf{D}$ and $\mathbf{E}$ be concrete categories with products. Objects of $\mathbf{D}$ and $\mathbf{E}$ are often just denoted by their underlying sets. A *schizophrenic object* is a set S which is the underlying set of an object S or $S_{\mathbf{D}}$ of $\mathbf{D}$ and an object S or $S_{\mathbf{E}}$ of $\mathbf{E}$. Now for an object D of $\mathbf{D}$, the morphism set $D^* = \mathbf{D}(D, S)$ is a subset of the product S^D. If this subset inherits the structure of $S_{\mathbf{E}}^D$ as a product object of $\mathbf{E}$, then the assignment $D \mapsto D^*$ may become the object part of a contravariant functor $F : \mathbf{D} \to \mathbf{E}$. Similarly, for an object E of $\mathbf{E}$, define $E^* = \mathbf{E}(E, S)$. The assignment $E \mapsto E^*$ may then become the object part of a contravariant functor $G : \mathbf{E} \to \mathbf{D}$. If the composites FG and GF are naturally isomorphic to the identity functors on their domains, then the categories $\mathbf{D}$ and $\mathbf{E}$ are said to be *dual* or *dually equivalent* via the schizophrenic object S.

Let T be the *one-dimensional torus*, the group S^1 of complex numbers of unit modulus under multiplication, or the quotient abelian group $\mathbb{R}/\mathbb{Z}$. The torus may be equipped with the subset topology induced from the Euclidean topology on $\mathbb{C}$, or the quotient topology induced from the Euclidean topology on $\mathbb{R}$. Then *Pontryagin duality* is the dual equivalence of the category $\mathbf{A}$ of locally compact abelian topological groups with itself that is given by T as schizophrenic object. For a locally compact abelian group A, the dual object A^* is called the *character group* of A. In particular, the compact group T is the character group of the discrete group $\mathbb{Z}$ of integers, while the locally compact group $\mathbb{R}$ of reals is its own character group.

Suppose that categories $\mathbf{D}$ and $\mathbf{E}$ are dually equivalent by a pair of contravariant functors F and G. Let J be a small category. Then objects of the functor category $\mathbf{D}^J$ are considered as *diagrams* in $\mathbf{D}$, images of the small category J. Similarly, objects of the functor category $\mathbf{E}^{J^{\mathrm{op}}}$ (consisting of covariant functors to $\mathbf{E}$ from the opposite J^{op} of J, or contravariant functors from J to $\mathbf{E}$) are images of the opposite category J^{op}. For an object $\theta : J \to \mathbf{D}$ of $\mathbf{D}^J$, the duality between $\mathbf{D}$ and $\mathbf{E}$ gives a *dual diagram* $\theta F : J \to \mathbf{E}$ in $\mathbf{E}^{J^{\mathrm{op}}}$. Similarly, a diagram $\varphi : J^{\mathrm{op}} \to \mathbf{E}$ in $\mathbf{E}$ yields a dual diagram $\varphi G : J \to \mathbf{D}$ in D. The double duals $\theta FG : J \to \mathbf{D}$ and $\varphi GF : J \to \mathbf{E}$ are naturally isomorphic to their respective primals θ and φ. Thus one obtains a *diagrammatic duality* between the functor categories $\mathbf{D}^J$ and $\mathbf{E}^{J^{\mathrm{op}}}$.

5. Duality for central piques

Let P be a locally compact central pique, with inner multiplication group H and pointed idempotent 0. The cloop $B(P)$ of P is a locally compact abelian group, with dual abelian group $B(P)^*$. Let H act from the left on $B(P)^*$ by $(b)^h\beta = (b^h)\beta$ for $h \in H$, $\beta \in B(P)^*$, and $b \in P$. A quasigroup operation is defined on $B(P)^*$ by

$$\xi \cdot \eta = {}^R\xi + {}^L\eta. \tag{5.5}$$

Definition 5.1. The pique P^* *dual* to P consists of the space $B(P)^*$ equipped with the quasigroup operation (5.5), and pointed by the trivial character 0 of $B(P)$.

Example 5.2. Consider the pique $(\mathbb{R}, -, 0)$ of real numbers under subtraction. It is a topological pique under the usual, locally compact (Euclidean) topology on $\mathbb{R}$. Since $xR(0) = x - 0 = x$ and $xL(0) = 0 - x = -x$, the $\langle R, L\rangle$-module structure on $\mathbb{R}$ is given by $xR = x$ and $xL = -x$ for each real number x. The character group of the cloop $B(\mathbb{R}, -, 0) = \mathbb{R}$ is again the usual abelian group $\mathbb{R}$. For a real number ξ considered as the character $\mathbb{R} \to S^1; x \mapsto \exp(2\pi i x\xi)$, the corresponding left actions are given by ${}^R\xi = \xi$ and ${}^L\xi = -\xi$. By (5.5), the dual pique $(\mathbb{R}, -, 0)^*$ has the subtraction $(\xi, \eta) \mapsto \xi - \eta$ as its quasigroup operation. Similar considerations show that the dual of the pique $(\mathbb{Z}, -, 0)$ of integers under subtraction is the unit circle $(S^1, \circ, 1)$ under $z \circ w = z\overline{w}$ for complex numbers z, w of unit modulus. Here z is considered as the character $\mathbb{Z} \to S^1; n \mapsto z^n$.

Let J be the free group on two generators R and L. Consider J as a small category with a single object $\circ$, and with morphisms corresponding to the elements of the group [2, Ch.1, §2]. The following proposition identifies locally compact central piques as J-diagrams in the category $\mathbf{A}$ of locally compact abelian groups.

Proposition 5.3. *The category of locally compact central piques is equivalent to the functor category* $\mathbf{A}^J$. *Under this equivalence, a pique* P *with cloop* $B(P)$ *determines the* covariant *functor* $J \to \mathbf{A}$ *specified uniquely by*

$$\theta : \begin{cases} R \mapsto \big(R : B(P) \to B(P)\big), \\ L \mapsto \big(L : B(P) \to B(P)\big). \end{cases} \tag{5.6}$$

Conversely, a functor $\varphi : J \to \mathbf{A}$ with $P = \circ\varphi$ determines a locally compact central pique on P with

$$x \cdot y = xR^{\varphi} + yL^{\varphi} \tag{5.7}$$

for x, y in P.

Proof. The functor (5.6) maps to $\mathbf{A}$, since the cloop of a locally compact central pique is a locally compact abelian group, and since the right and left multiplications by the pointed idempotent are homeomorphic homomorphisms. Conversely, (5.7) defines a locally compact central pique, since the automorphisms R^{φ}, L^{φ}, and their inverses are continuous. □

Proposition 5.3 has a dual counterpart.

Proposition 5.4. *The category of locally compact central piques is equivalent to the functor category $\mathbf{A}^{J^{\mathrm{op}}}$. Under this equivalence, a pique P' with cloop $B(P')$ determines the* contravariant *functor $J \to \mathbf{A}$ specified uniquely by*

$$\theta : \begin{cases} R \mapsto \big(R : B(P') \to B(P')\big), \\ L \mapsto \big(L : B(P') \to B(P')\big). \end{cases} \tag{5.8}$$

Conversely, a contravariant functor $\varphi : J \to \mathbf{A}$ with $P' = \circ\varphi$ determines a locally compact central pique on P' with

$$x \cdot y = R^{\varphi}x + L^{\varphi}y \tag{5.9}$$

for x, y in P'. In particular, if a central pique P corresponds to the functor (5.6), then the dual pique P^* corresponds to the contravariant functor

$$\theta^* : \begin{cases} R \mapsto \big(R^* : B(P)^* \to B(P)^*\big), \\ L \mapsto \big(L^* : B(P)^* \to B(P)^*\big). \end{cases} \tag{5.10}$$

Proof. If the primal pique P is given by the functor (5.6) according to Proposition 5.3, one obtains $B(P)$ as a right $\langle R, L\rangle$ module according to Proposition 3.4. Diagrammatic duality then gives the dual group $B(P)^*$ as the left $\langle R, L\rangle$-module corresponding to the dual pique P^* of Definition 5.1, equivalent to the functor (5.10). □

Theorem 5.5. *The correspondence between the covariant functors θ of (5.6) and the contravariant functors θ^* of (5.10) yields a duality for the category of locally compact central piques.*

Proof. If P corresponds under the equivalence of Proposition 5.3 to the functor (5.6), then P^* corresponds to (5.10). In turn, P^{**} corresponds to the dual of this latter functor, namely the covariant functor θ^{**} with

$$\theta^{**} : \begin{cases} R \mapsto \left(R^{**} : B(P)^{**} \to B(P)^{**}\right), \\ L \mapsto \left(L^{**} : B(P)^{**} \to B(P)^{**}\right). \end{cases}$$

Since θ^{**} is naturally isomorphic to θ, the double dual pique P^{**} is naturally isomorphic to the primal P. □

6. Suvorov duality

An algebra is said to be *idempotent* if each singleton forms a subalgebra. It is said to be *entropic* if each operation is a homomorphism. Finally, it is said to be a *mode* if it is both idempotent and entropic [3]. For quasigroups, idempotence reduces to satisfaction of the identity

$$x \cdot x = x \tag{6.11}$$

[7, (3)], while entropicity reduces to satisfaction of the identity

$$xy \cdot zt = xz \cdot yt \tag{6.12}$$

[7, (2)]. Comparing with (3.2), it is apparent that a central pique P is a mode if and only if

$$R + L = 1 \tag{6.13}$$

and

$$RL = LR \tag{6.14}$$

in the endomorphism ring of the cloop $B(P)$. In fact P is a mode if and only if (6.13) alone is satisfied, since then $RL = R(1 - R) = (1 - R)R = LR$. Conversely, any quasigroup mode P is central, since the diagonal $\widehat{P}$ is the preimage of the singleton subquasigroup $\{\widehat{P}\}$ under the quasigroup homomorphism

$$P^2 \to P^2/\widehat{P}; \quad (x, y) \mapsto \widehat{P}(x, y)$$

to the set $\{\widehat{P}(x, y) \mid x, y \in P\}$ of cosets of $\widehat{P}$ under the multiplication

$$\widehat{P}(x, y) \cdot \widehat{P}(x', y') = \widehat{P}\widehat{P} \cdot \big((x, y)(x', y')\big) = \widehat{P}(xx', yy')$$

well-defined by the entropic law (6.12).

Suvarov duality is now recovered as follows.

Corollary 6.1. *The duality of Theorem 5.5 restricts to a self-duality for the category of locally compact quasigroup modes.*

Proof. If the image of the functor (5.6) satisfies (6.13), then so does the image of the dual functor (5.10). Since the images of the morphism parts of the functors are commutative groups, there is no distinction between covariant and contravariant functors. □

References

[1] **O. Chein et al**: *Quasigroups and Loops: Theory and Applications*, Heldermann, Berlin, 1990.

[2] **S. Mac Lane**: *Categories for the Working Mathematician*, Springer-Verlag, Berlin, 1971.

[3] **A. B. Romanowska and J. D. H. Smith**: *Modes*, World Scientific, Singapore, 2002.

[4] **J. D. H. Smith**: *Characters of central piques*, J. Algebra **279** (2004), 437 – 450.

[5] **J. D. H. Smith**: *An Introduction to Quasigroups and their Representations*, CRC Press, Boca Raton, FL, to appear.

[6] **J. D. H. Smith and A. B. Romanowska**: *Post-Modern Algebra*, Wiley, New York, NY, 1999.

[7] **N. M. Suvarov**: *A duality theorem for distributive transitive topological quasigroups* (Russian), Matem. Issled. **4** (1969), 162 – 166.

Received October 31, 2005

A. B. Romanowska
Faculty of Mathematics and Information Sciences
Warsaw University of Technology
00-661 Warsaw
Poland
e-mail: aroman@alpha.mini.pw.edu.pl

J. D. H. Smith
Department of Mathematics
Iowa State University
Ames, Iowa 50011
U.S.A
e-mail: jdhsmith@iastate.edu, url: http://www.math.iastate.edu/jdhsmith/

Quasigroups and Related Systems **14** (2006), 91 – 109

Loops in Relativistic Dynamics

Tzvi Scarr

Abstract

The Einstein velocity addition loop and the symmetric velocity addition loop are used to develop relativistic dynamic equations. Since these loops are highly non-commutative, the question arises whether one should use the left or the right translations of these loops. We show that while the left translations are well-suited to relativistic dynamics, the right translations are problematic. We hypothesize that using the left translations is equivalent to a generalized form of the Equivalence Principle.

1. Introduction

This paper is about two loops which play a central role in Special Relativity. The first is the Einstein velocity addition loop $(D_v, \oplus_E)$, where

$$D_v = \{\mathbf{v} \in \mathbb{R}^3 : |\mathbf{v}| < c\}$$

(c = the speed of light) is the ball of relativistically admissible velocities and $\mathbf{v} \oplus_E \mathbf{u}$ is the relativistic sum of the two velocities $\mathbf{v}$ and $\mathbf{u}$. This is a left Bruck loop. The loop operation $\oplus_E$ is constructed from the Lorentz transformations between two inertial systems. This construction will be carried out in Section 2. Einstein velocity addition is, in general, not commutative. In fact, $\mathbf{v} \oplus_E \mathbf{u} = \mathbf{u} \oplus_E \mathbf{v}$ if and only if $\mathbf{v}$ and $\mathbf{u}$ are parallel.

The second loop under investigation involves a new dynamic variable, called *symmetric velocity*, defined as follows. If the relative velocity between two inertial systems is $\mathbf{v}$, then the symmetric velocity between the systems is the unique velocity $\mathbf{w}$ such that $\mathbf{w} \oplus_E \mathbf{w} = \mathbf{v}$. Thus the symmetric velocity is the *relativistic half* of the given velocity. Let $D_s = \{\mathbf{v} \in \mathbb{R}^3 : |\mathbf{v}| < 1\}$

2000 Mathematics Subject Classification: 20N05, 83A05
Keywords: loops, special relativity, relativistic dynamics, Einstein velocity addition, symmetric velocity, Equivalence Principle

denote the set of relativistically admissible *symmetric* velocities (normalized to $c = 1$). The set D_s admits a binary operation $\oplus_s$, the addition of symmetric velocities, which makes $(D_s, \oplus_s)$ a loop.

The two loops $(D_v, \oplus_E)$ and $(D_s, \oplus_s)$ are isotopic as *topological loops*. Indeed, the function $\Psi : D_v \to D_s$ which maps a given velocity $\mathbf{v}$ to its corresponding symmetric velocity $\mathbf{w}$ is a homeomorphism which also respects the loop operations:

$$\Psi(\mathbf{v} \oplus_E \mathbf{u}) = \Psi(\mathbf{v}) \oplus_s \Psi(\mathbf{u}). \tag{1}$$

See Section 5 for explicit definitions of Ψ and Ψ^{-1}.

Despite the above isotopy, these loops *are* different. $(D_v, \oplus_E)$ is a left Bruck loop, whereas $(D_s, \oplus_s)$ is not. Moreover, these two loops behave differently *geometrically*. Friedman and Semon [2] have already exploited this difference. They used symmetric velocity and obtained an *analytic* solution for the motion of an electric charge in a uniform, constant electromagnetic field $\mathbf{E}, \mathbf{B}$ in which $\mathbf{E}$ and $\mathbf{B}$ are perpendicular. The first *explicit* solution to this problem was found in 2002 by Takeuchi [5].

The left translations of $(D_v, \oplus_E)$ (respectively, $(D_s, \oplus_s)$) generate a group of automorphisms of D_v (respectively, D_s). In turns out that in the case of $(D_v, \oplus_E)$, the automorphisms are *projective* (also called *affine*). This means that line segments are mapped to line segments. In this way, D_v can be seen as a subset of *projective space* $\mathbb{P}_3$. In contrast, the automorphisms induced by *symmetric* velocity are *conformal*. Thus while the two automorphism groups are isomorphic as groups, they are quite different *geometrically*.

The use of these two loops in developing relativistic dynamics is new and brings with it an interesting *dilemma*. Relativistic dynamics is concerned with describing the motion of an object whose velocity is changing with time due to a force. Since the velocities are bounded by c, they must be added relativistically. Over an infinitesimal time period dt, the force adds a change $d\mathbf{v}$ to the velocity $\mathbf{v}$. The new velocity will be $\mathbf{v} \oplus_E d\mathbf{v}$. Thus, velocity addition lies at the heart of relativistic dynamics, and it is natural to use the loop $(D_v, \oplus_E)$ to develop relativistic dynamics.

Now comes the dilemma. Is the new velocity really $\mathbf{v} \oplus_E d\mathbf{v}$? Or is it $d\mathbf{v} \oplus_E \mathbf{v}$? Since Einstein velocity addition is, in general, not commutative, we must choose between having the force act on the left or on the right. At first glance, this choice seems arbitrary. There is no *a priori* preference. Why should we prefer one over the other? And how does the *force* know which

side to act on? Furthermore, does it matter? Does the dynamics based on left translations coincide with the dynamics based on right translations? The answer to this last question will be interesting either way. Agreement of "left" and "right" dynamics would be fascinating given the highly non-commutative nature of the velocity addition. On the other hand, if the two dynamics are at odds, we will then be faced with two additional questions: *Which* dynamics does nature use? *Why* does nature use this one?

Unfortunately, we cannot yet compare "left" and "right" dynamics because no one to date has succeeded in using the right translations to develop relativistic dynamics. Indeed, Friedman [1] uses the *left* translations of the Einstein velocity addition loop $(D_v, \oplus_E)$ to derive the relative dynamics equation

$$m_0 \frac{d\mathbf{v}(\tau)}{d\tau} = q(\mathbf{E} + \mathbf{v}(\tau) \times \mathbf{B} - c^{-2}\langle \mathbf{v}(\tau)|\mathbf{E}\rangle \mathbf{v}(\tau)) \tag{2}$$

for a particle of charge q and rest-mass m_0 in an electromagnetic field $\mathbf{E}, \mathbf{B}$. Here, τ is the proper time of the particle. Friedman's development is straightforward. The *right* translations, on the other hand, possess some inherent pathologies. We will attempt to explain this asymmetry in terms of the physical interpretation of the loop operations.

Note that the traditional approach to relativistic dynamics does not encounter the above dilemma. In fact, relativistic dynamics is usually developed without reference to Einstein velocity addition at all. In [3], for example, one starts with the assumption that the force on, say, a charged particle is equal to the rate of change of the particle's relativistic momentum. Since a particle with charge q and velocity $\mathbf{v}$ in an electromagnetic field $\mathbf{E}, \mathbf{B}$ experiences a force $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, the relativistic dynamics equation in this case is

$$m_0 \frac{d(\gamma \mathbf{v})}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{3}$$

In [1], it is shown that (2) and (3) are equivalent. Note that although the traditional approach avoids our dilemma, the equivalence of (2) and (3) means that the traditional approach implicitly assumes that the force acts on the *left*. See also [4].

This paper is organized as follows. In the next section, we construct the Einstein velocity addition loop from the Lorentz transformations. In Section 3, the left translations of this loop are used to derive the relativistic dynamics equation (2). Section 4 describes the difficulties inherent in using the right translations to develop relativistic dynamics. Section 5 is devoted

to the symmetric velocity addition loop. Here, also, we will see that the left translations are preferred over the right. In Section 6, we discuss possible reasons why the left and right translations should behave so differently. The final section offers suggestions for further research. One direction is to develop relativistic dynamics using the triple product to overcome the difficulties of the right translations. Another approach is to show that using the left translations is actually equivalent to the Equivalence Principle. The latter idea will be taken up in a forthcoming paper.

2. Construction of the Einstein velocity addition loop

In this section, we will construct the Einstein velocity addition loop from the Lorentz spacetime transformation between two inertial systems K and K'. We assume that the spatial axes of K are parallel to those of K' and that at time $t = 0$, the origins of the two systems coincided. The spacetime coordinates of an event in K will be denoted by $\begin{pmatrix} t \\ \mathbf{r} \end{pmatrix}$, where $t \in \mathbb{R}$ is the time of the event and $\mathbf{r} \in \mathbb{R}^3$ represents the location of the event. The coordinates of the same event in K' will be denoted by $\begin{pmatrix} t' \\ \mathbf{r}' \end{pmatrix}$.

Suppose that the velocity of K' with respect to K is $\mathbf{v}$. Then the Lorentz transformation from K' to K is

$$\begin{pmatrix} t \\ \mathbf{r} \end{pmatrix} = \gamma \begin{pmatrix} 1 & c^{-2}\mathbf{v}^T \\ \mathbf{v} & P_{\mathbf{v}} + \alpha(I - P_{\mathbf{v}}) \end{pmatrix} \begin{pmatrix} t' \\ \mathbf{r}' \end{pmatrix}, \tag{4}$$

where $\gamma = \gamma(\mathbf{v}) = \frac{1}{\sqrt{1-\frac{|\mathbf{v}|^2}{c^2}}}$, $\alpha = \alpha(\mathbf{v}) = \frac{1}{\gamma(\mathbf{v})}$, $\mathbf{v}^T$ denotes the transpose of $\mathbf{v}$, and $P_{\mathbf{v}}$ denotes the projection operator onto $\mathbf{v}$.

The physical definition of the Einstein velocity addition is as follows. We are given that the velocity of K' with respect to K is $\mathbf{v}$. Suppose that an observer at rest in system K' measures an object's velocity as $\mathbf{u}$. Then the velocity of this object as measured by an observer at rest in system K is called the *relativistic sum* of $\mathbf{v}$ and $\mathbf{u}$ and is denoted by $\mathbf{v} \oplus_E \mathbf{u}$.

Consider motion with uniform velocity $\mathbf{u}$ in system K'. The world-line of this motion is $\begin{pmatrix} t' \\ \mathbf{u}t' \end{pmatrix}$. From (4), this world-line in system K is

$$\gamma \begin{pmatrix} t' + \frac{\mathbf{v}^T\mathbf{u}t'}{c^2} \\ \mathbf{v}t' + t'P_{\mathbf{v}}\mathbf{u} + \alpha t'(I - P_{\mathbf{v}})\mathbf{u} \end{pmatrix} \tag{5}$$

or

$$\gamma t' \begin{pmatrix} 1 + \frac{\langle \mathbf{v}|\mathbf{u}\rangle}{c^2} \\ \mathbf{v} + \mathbf{u}_{\|} + \alpha \mathbf{u}_{\perp} \end{pmatrix}, \tag{6}$$

where $\mathbf{u}_{\|} = P_{\mathbf{v}}\mathbf{u}$ denotes the component of $\mathbf{u}$ parallel to $\mathbf{v}$ and $\mathbf{u}_{\perp} = (I - P_{\mathbf{v}})\mathbf{u}$ denotes the component of $\mathbf{u}$ perpendicular to $\mathbf{v}$. This defines a uniform motion in system K with velocity

$$\mathbf{v} \oplus_E \mathbf{u} = \frac{\mathbf{v} + \mathbf{u}_{\|} + \alpha \mathbf{u}_{\perp}}{1 + \frac{\langle \mathbf{v}|\mathbf{u}\rangle}{c^2}}, \tag{7}$$

with $\alpha = \alpha(\mathbf{v}) = \sqrt{1 - \frac{|\mathbf{v}|^2}{c^2}}$. This defines the binary operation $\oplus_E$ on D_v. The pair $(D_v, \oplus_E)$ is a left Bruck loop.

In case $\mathbf{v}$ and $\mathbf{u}$ are parallel, the Einstein velocity addition reduces to

$$\mathbf{v} \oplus_E \mathbf{u} = \frac{\mathbf{v} + \mathbf{u}}{1 + \frac{vu}{c^2}}, \tag{8}$$

where $v = |\mathbf{v}|$ and $u = |\mathbf{u}|$. In case $\mathbf{v}$ and $\mathbf{u}$ are perpendicular, the formula becomes

$$\mathbf{v} \oplus_E \mathbf{u} = \mathbf{v} + \alpha(\mathbf{v})\mathbf{u}. \tag{9}$$

Note that the velocity addition is commutative only for parallel velocities.

3. Left translations

In [1], the left translations of the loop $(D_v, \oplus_E)$ are used to obtain the relativistic dynamics equation

$$m_0 \frac{d\mathbf{v}(\tau)}{d\tau} = q(\mathbf{E} + \mathbf{v}(\tau) \times \mathbf{B} - c^{-2}\langle \mathbf{v}(\tau)|\mathbf{E}\rangle \mathbf{v}(\tau)), \tag{10}$$

where τ is the proper time of the particle. It is then shown that (10) is equivalent to (3). Here we give an outline of that development. For details, see [1].

For each $\mathbf{v}$ in the velocity ball D_v, we define the left translation $L_{\mathbf{v}} : D_v \to D_v$ by

$$L_{\mathbf{v}}(\mathbf{u}) = \mathbf{v} \oplus_E \mathbf{u}. \tag{11}$$

See Figure 1.

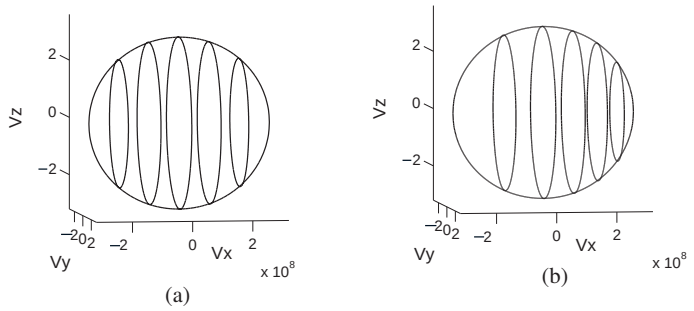

Figure 1. *Action of the velocity addition on* D_v.
(a) A set of 5 uniformly spaced discs Δ_j obtained by intersecting the three-dimensional velocity ball D_v of radius $c = 3{\cdot}10^8$m/s with y-z planes at $x = 0, \pm 10^8, \pm 2{\cdot}10^8$m/s. ($b$) The images of these Δ_j under the left translation $L_{\mathbf{v}}(\mathbf{u}) = \mathbf{v} \oplus_E \mathbf{u}$, with $\mathbf{v} = (10^8, 0, 0)$m/s. Note that $L_{\mathbf{v}}(\Delta_j)$ is also a disc in D_v, perpendicular to $\mathbf{v}$ and moved in the direction of $\mathbf{v}$. On each disc Δ_j, the map $L_{\mathbf{v}}$ acts as multiplication by a constant in the component of $\mathbf{u}$ perpendicular to $\mathbf{v}$.

The left translations have some nice properties. First, each left translation $L_{\mathbf{v}}$ is a *projective* automorphism of D_v. To appreciate the projective geometry at work here, envision the action of $L_{\mathbf{v}}$ on D_v as follows. Fix a velocity $\mathbf{u} \in D_v$. Identify $\mathbf{u}$ with the intersection of the world-line $L = \begin{pmatrix} t \\ \mathbf{u}t \end{pmatrix}$ in the inertial system K and the plane $\Pi = \{(1, \mathbf{r}) : \mathbf{r} \in \mathbb{R}\}$. Let K' be an inertial system moving with relative velocity $\mathbf{v}$ with respect to K. Applying the Lorentz transformation from K to K' to the line L yields a line L' in K' whose intersection with Π is $\mathbf{v} \oplus_E \mathbf{u}$.

The second nice property is closure under inverses. In fact $L_{\mathbf{v}}^{-1} = L_{-\mathbf{v}}$. The above two properties combine to make the following useful characterization of the group $Aut_p(D_v)$ of all projective automorphisms of D_v. Let ψ be any projective automorphism of D_v. Set $\mathbf{v} = \psi(0)$ and $U = L_{\mathbf{v}}^{-1}\psi$. Then U is a projective map that maps $0 \to 0$ and is thus a linear map which can be represented by a 3×3 matrix. Since U maps D_v onto itself, it is an isometry and is represented by an orthogonal matrix. Since $\psi = L_{\mathbf{v}}U$, the group $Aut_p(D_v)$ is defined by

$$Aut_p(D_v) = \{L_{\mathbf{v}}U : \mathbf{v} \in D_v,\ U \in O(3)\}. \tag{12}$$

We write $L_{\mathbf{v},U}$ instead of $L_{\mathbf{v}}U$.

The group $Aut_p(D_v)$ is a representation of the Lorentz group by affine maps. It is a real Lie group of dimension 6, since any element of the group

is determined by an element $\mathbf{v}$ of the three-dimensional open ball of radius c in $\mathbb{R}^3$ and an element U of the three-dimensional orthogonal group $O(3)$.

The dynamics equation (10) will be constructed from the elements of the Lie algebra $aut_p(D_v)$ of $Aut_p(D_v)$. The elements of a Lie algebra are, by definition, the tangent space of the identity of the corresponding Lie group. To obtain the elements of $aut_p(D_v)$, let $g(s)$ be a differentiable curve from a neighborhood I_0 of 0 into $Aut_p(D_v)$, with $g(0) = L_{\mathbf{0},I}$, the identity of $Aut_p(D_v)$. Then $g(s)$ has the form

$$g(s) = L_{\mathbf{v}(s),U(s)}, \tag{13}$$

where $\mathbf{v} : I_0 \to D_v$ is a differentiable function satisfying $\mathbf{v}(0) = \mathbf{0}$ and $U(s) : I_0 \to O(3)$ is differentiable and satisfies $U(0) = I$.

For an example of such a curve $g(s)$, fix $\mathbf{v} \in D_v$. Let $\mathbf{j} = \mathbf{v}/|\mathbf{v}|$ and define $k = \tanh^{-1}(|\mathbf{v}|/c)$. For $s \in \mathbb{R}$, define $\mathbf{b}(s) = \tanh(sk)c\mathbf{j}$. The resulting curve $g(s) := L_{\mathbf{b}(s)}$ is called the *one-parameter subgroup generated by* $L_{\mathbf{v}}$. See Figure 2.

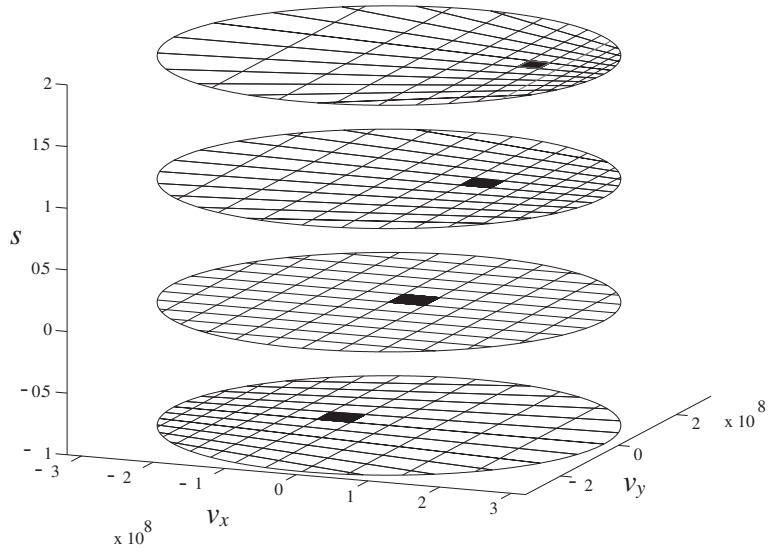

Figure 2. Action of a one-parameter subgroup on D_v.

The effect on a two-dimensional section of D_v by the one-parameter subgroup $g(s)$ generated by the map $L_{\mathbf{v}}$, for $s = -1, 0, 1, 2$. One cell of the grid has been darkened along with its images to help visualize the effect of the transformation. Note that $g(-1) = L_{\mathbf{v}}^{-1} = L_{-\mathbf{v}}$, $g(0) = I$-the identity, $g(1) = L_{\mathbf{v}}$ and $g(2) = L_{\mathbf{v}}^2 = L_{\mathbf{v}\oplus_E\mathbf{v}}$.

We denote by δ the element of $aut_p(D_v)$ generated by $g(s)$. For any fixed $\mathbf{u} \in D_v$, $g(s)(\mathbf{u})$ is a smooth curve in D_v, with $g(0)(\mathbf{u}) = \mathbf{u}$, and $\delta(\mathbf{u})$ is a tangent vector to this line. Thus, the elements of $aut_p(D_v)$ are vector fields $\delta(\mathbf{u})$ on D_v defined by

$$\delta(\mathbf{u}) = \frac{d}{ds} g(s)(\mathbf{u})\Big|_{s=0}. \tag{14}$$

Note that (14) is equivalent to using $d\mathbf{v} \oplus_E \mathbf{v}$ and *not* $\mathbf{v} \oplus_E d\mathbf{v}$ for the velocity at time $t + dt$.

The explicit form of $\delta(\mathbf{u})$ is calculated in [1]. There it is shown that the Lie algebra

$$aut_p(D_v) = \{\delta_{\mathbf{E},\mathbf{B}} : \mathbf{E}, \mathbf{B} \in \mathbb{R}^3\}, \tag{15}$$

where $\delta_{\mathbf{E},\mathbf{B}} : D_v \to \mathbb{R}^3$ is the vector field defined by

$$\delta_{\mathbf{E},\mathbf{B}}(\mathbf{u}) = \mathbf{E} + \mathbf{u} \times \mathbf{B} - c^{-2}\langle \mathbf{u} \,|\mathbf{E}\rangle \mathbf{u}. \tag{16}$$

Note that each generator $\delta_{\mathbf{E},\mathbf{B}}(\mathbf{u})$ is a second-degree polynomial in $\mathbf{u}$. The quadratic term can be used to derive the *triple product* associated with D_v as a *Bounded Symmetric Domain*. Moreover, these generators give the correct formulas for the transformation of an electromagnetic field between two inertial systems. Two examples of these vector fields are shown in Figures 3 and 4.

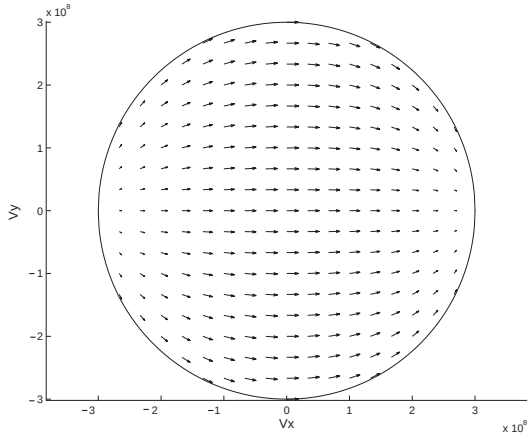

Figure 3. *The vector field generated by an electric field* $\mathbf{E}$.
The vector field $q/m \cdot \delta_{\mathbf{E},\mathbf{B}}$ on a two-dimensional section of D_v, with $q/m = 10^7 C/kg$, $\mathbf{E} = (2, 0, 0)V/m$ and $\mathbf{B} = 0$. Since $\mathbf{E}$ is in the positive direction of the v_x-axis, the

field tends to move particles in this direction. However, near the edge of D_v, the vectors either shrink to zero magnitude or become nearly tangent to D_v, reflecting the fact that the flow generated by this field cannot leave D_v.

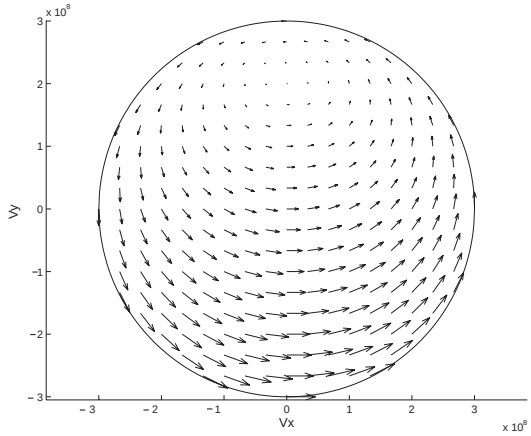

Figure 4. *The vector field generated by an electromagnetic field* $\mathbf{E}, \mathbf{B}$. The vector field $q/m \cdot \delta_{\mathbf{E},\mathbf{B}}$ on a two-dimensional section of D_v, with $q/m = 10^7 C/kg$, $\mathbf{E} = (2,0,0)V/m$ and $c\mathbf{B} = (0,0,3)V/m$. Here, the addition of a magnetic field $\mathbf{B}$ causes a rotation.

Using the generator $\delta_{\mathbf{E},\mathbf{B}} \in aut_p(D_v)$ defined by (16) to represent the force on a particle with charge q and rest-mass m_0, we obtain the relativistic dynamics equation

$$\frac{d\mathbf{v}(\tau)}{d\tau} = \frac{q}{m_0}\delta_{\mathbf{E},\mathbf{B}}(\mathbf{v}(\tau))$$

or

$$m_0\frac{d\mathbf{v}(\tau)}{d\tau} = q(\mathbf{E} + \mathbf{v}(\tau) \times \mathbf{B} - c^{-2}\langle \mathbf{v}(\tau)|\mathbf{E}\rangle \mathbf{v}(\tau)), \tag{17}$$

where τ is the proper time of the particle. It is shown in [1] that (17) is equivalent to (3).

4. Right translations

When we try to mimic the development of the previous section using the right translations, we run into problems.

For each $\mathbf{v} \in D_v$, we define the right translation $R_{\mathbf{v}} : D_v \to D_v$ by

$$R_{\mathbf{v}}(\mathbf{u}) = \mathbf{u} \oplus_E \mathbf{v}. \tag{18}$$

Unfortunately, the right translations do not possess any of the nice properties of the left translations. The map $R_{\mathbf{v}}$ is not projective. It's not even analytic. Moreover,

$$R_{\mathbf{v}}^{-1} \neq R_{-\mathbf{v}}. \tag{19}$$

In fact, $R_{\mathbf{v}}^{-1}$ is not a right translation at all. We will express $R_{\mathbf{v}}^{-1}(\mathbf{u})$ using Ungar's gyration operator [6]. For $\mathbf{x}, \mathbf{y} \in D_v$, define $\mathrm{gyr}[\mathbf{x}, \mathbf{y}] : D_v \to D_v$ by

$$\mathrm{gyr}[\mathbf{x}, \mathbf{y}](\mathbf{z}) = -(\mathbf{x} \oplus_E \mathbf{y}) \oplus_E (\mathbf{x} \oplus_E (\mathbf{y} \oplus_E \mathbf{z})). \tag{20}$$

Then

$$R_{\mathbf{v}}^{-1}(\mathbf{u}) = \mathbf{u} \oplus_E -\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{v}. \tag{21}$$

This last equation shows that $R_{\mathbf{v}}^{-1}$ is not a right translation. It is not clear at all how to proceed from this point in developing relativistic dynamics. We think the difficulties might be overcome by using the triple product associated with D_v as a Bounded Symmetric Domain, but this approach is still in the beginning stages. See Section 7.

5. Symmetric velocity addition

In this section, we construct the loop $(D_s, \oplus_s)$ of symmetric velocities. We derive the formula for the addition of symmetric velocities from the physical definition of this addition. The left translations of $(D_s, \oplus_s)$, which belong to the group $Aut_c(D_s)$ of all *conformal* automorphisms of D_s, are then used to derive the relativistic dynamics equation for symmetric velocities. The elements of the Lie algebra $aut_c(D_s)$ will be expressed in terms of a *triple product*. We also obtain a very useful two-dimensional version of $(D_s, \oplus_s)$. This version is usually simpler to work with and yet captures all of the properties of the full three-dimensional version. Here too, in the case of symmetric velocity, we will see that while the *left* translations yield a nice development of relativistic dynamics, the *right* translations are rather problematic, even in the simpler, two-dimensional case.

The definition of symmetric velocity is as follows. If the relative velocity between two inertial systems is $\mathbf{v}$, then the symmetric velocity between the systems is the unique velocity $\mathbf{w}_1$ such that

$$\mathbf{v} = \mathbf{w}_1 \oplus_E \mathbf{w}_1 = \frac{\mathbf{w}_1 + \mathbf{w}_1}{1 + \frac{|\mathbf{w}_1|}{c}\frac{|\mathbf{w}_1|}{c}} = \frac{2\mathbf{w}_1}{1 + |\mathbf{w}_1|^2/c^2}.$$

Instead of $\mathbf{w}_1$, we use a dimensionless vector $\mathbf{w} = \mathbf{w}_1/c$ and call it *s-velocity*. Thus, the relationship between an s-velocity $\mathbf{w}$ and its corresponding velocity $\mathbf{v}$ is given by the two formulas

$$\mathbf{w} = \Psi(\mathbf{v}) = \frac{\mathbf{v}/c}{1 + \sqrt{1 - |\mathbf{v}|^2/c^2}} \tag{22}$$

and

$$\mathbf{v} = \Psi^{-1}(\mathbf{w}) = \frac{2c\mathbf{w}}{1 + |\mathbf{w}|^2}. \tag{23}$$

The set of all relativistically admissible s-velocities form a unit ball

$$D_s = \{\mathbf{w} \in \mathbb{R}^3 : \ |\mathbf{w}| < 1\}. \tag{24}$$

The physical meaning of symmetric velocity is as follows. Consider two inertial systems with relative velocity $\mathbf{v}$ between them. Place two objects of equal mass (test masses) at the origin of each inertial system. The center of mass of the two objects will be called the *center of the two inertial systems*. The symmetric velocity is the velocity of each system with respect to the center of the systems, and the s-velocity is the dimensionless velocity of the systems with respect to their center (see Figure 5).

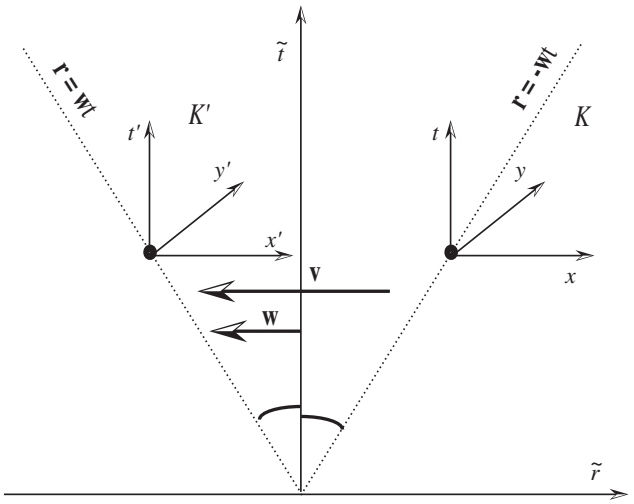

Figure 5. *The physical meaning of symmetric velocity.*
The physical meaning of symmetric velocity. Two inertial systems K and K' with relative velocity $\mathbf{v}$ between them are viewed from the system connected to their center. In this system, K and K' are each moving with velocity $\pm\mathbf{w}$.

The physical definition of s-velocity addition is as follows. Consider three inertial systems K_1, K_2 and K_3. We choose the space axes of K_2 to be parallel to the axes of K_1 and the axes of K_3 to be parallel to those of K_2. Denote their origins by O_1, O_2 and O_3, respectively. Denote by $\mathbf{a}$ the s-velocity of system K_2 with respect to K_1 and by $\mathbf{w}$ the s-velocity of system K_3 with respect to K_2. Then the s-velocity $\mathbf{w}_3$ of system K_3 with respect to K_1 (*i.e.*, the velocity of K_3 with respect to the center of systems K_1 and K_3) is called the *sum of the s-velocities* $\mathbf{a}$ and $\mathbf{w}$ and is denoted by $\mathbf{a} \oplus_s \mathbf{w}$ (see Figure 6). In other words, if $c\mathbf{a} \oplus_E c\mathbf{a} = \mathbf{v}$ and $c\mathbf{w} \oplus_E c\mathbf{w} = \mathbf{u}$, then $\mathbf{a} \oplus_s \mathbf{w}$ is $1/c$ times the relativistic half of $\mathbf{v} \oplus_E \mathbf{u}$.

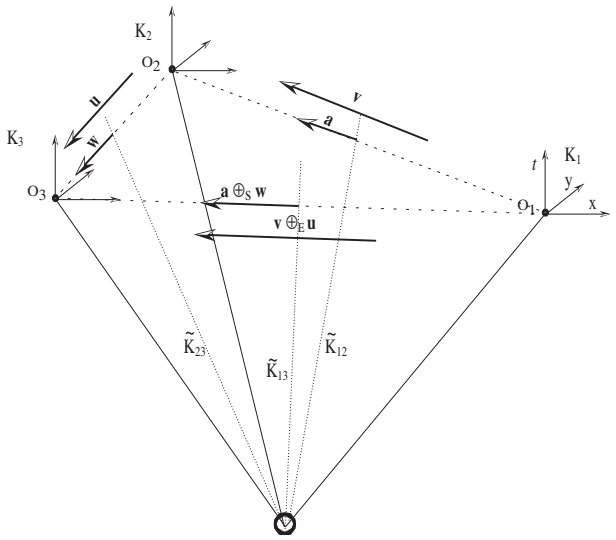

Figure 6. *The sum of s-velocities.*

The sum of s-velocities. Inertial systems K_1, K_2 and K_3, with origins O_1, O_2 and O_3, respectively, had a common origin at time $t = 0$. The line $\widetilde{K}_{12}$ is the world-line of the center of the two inertial systems K_1 and K_2. Similarly, the lines $\widetilde{K}_{23}$ and $\widetilde{K}_{13}$ represent the world-lines of the centers of the systems K_2, K_3 and K_1, K_3, respectively. The velocity of system K_2 with respect to system K_1 is $\mathbf{v}$, and its s-velocity $\mathbf{a}$ is the velocity of K_2 with respect to $\widetilde{K}_{12}$. Similarly, the velocity of system K_3 with respect to system K_2 is $\mathbf{u}$, and its s-velocity $\mathbf{w}$ is the velocity of K_3 with respect to $\widetilde{K}_{23}$. By definition of Einstein velocity addition, the velocity of system K_3 with respect to system K_1 is $\mathbf{v} \oplus_E \mathbf{u}$. The s-velocity of K_3 with respect to K_1, meaning the dimensionless velocity of K_3 with respect to $\widetilde{K}_{13}$, is called the sum of symmetric velocities $\mathbf{a}$ and $\mathbf{w}$ and is denoted by $\mathbf{a} \oplus_s \mathbf{w}$.

Using the above definition and formula (7) for Einstein velocity addition, we obtain the s-velocity-addition formula:

$$\mathbf{a} \oplus_s \mathbf{w} = \frac{(1 + |\mathbf{w}|^2 + 2\langle \mathbf{a} \mid \mathbf{w} \rangle)\mathbf{a} + (1 - |\mathbf{a}|^2)\mathbf{w}}{1 + |\mathbf{a}|^2|\mathbf{w}|^2 + 2\langle \mathbf{a} \mid \mathbf{w} \rangle}. \tag{25}$$

As in the case of Einstein velocity addition, it can be shown that $\mathbf{a} \oplus_s \mathbf{w} = \mathbf{w} \oplus_s \mathbf{a}$ if and only if $\mathbf{a}$ and $\mathbf{w}$ are parallel.

Note that $\mathbf{a} \oplus_s \mathbf{w}$ is a linear combination of $\mathbf{a}$ and $\mathbf{w}$ and therefore belongs to the plane Π generated by $\mathbf{a}$ and $\mathbf{w}$. This allows us to obtain a two-dimensional version of s-velocity addition. It is often sufficient (and easier!) to work with the two-dimensional version. Moreover, we obtain a new method of solving relativistic dynamic equations. If the motion under investigation has an invariant plane, then the relativistic dynamic equation for the symmetric velocity becomes a first-order *analytic* differential equation in one complex variable.

We obtain the two-dimensional version of s-velocity addition by imposing a complex structure on the plane Π. In other words, we treat the disk $\Delta = D_s \cap \Pi$ as a copy of the unit disk $|z| < 1$ of the complex plane $\mathbb{C}$. Denote by a the complex number corresponding to the vector $\mathbf{a}$ and by w the complex number corresponding to the vector $\mathbf{w}$. We use the identities

$$Re\,\langle a \mid w \rangle = \frac{\overline{a}w + a\overline{w}}{2},\ |w|^2 = w\overline{w}, \tag{26}$$

where the bar denotes complex conjugation, to convert (25) into our two-dimensional version:

$$a \oplus_s w = \frac{(1 + w\overline{w} + \overline{a}w + a\overline{w})a + (1 - a\overline{a})w}{1 + a\overline{a}w\overline{w} + \overline{a}w + a\overline{w}} \tag{27}$$

$$= \frac{(a + w)(1 + a\overline{w})}{(1 + \overline{a}w)(1 + a\overline{w})} = \frac{a + w}{1 + \overline{a}w}. \tag{28}$$

This is the well-known Möbius transformation of the complex unit disk. Thus, s-velocity addition is a generalization of the Möbius addition of complex numbers (see Figure 7).

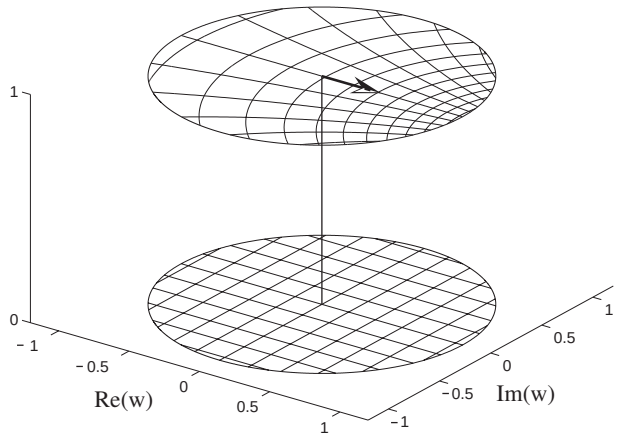

Figure 7. Symmetric velocity addition on D_s.
Symmetric velocity addition $a \oplus_s w$ for $a = 0.4$. The lower circle in the figure is the unit disc of the complex plane, representing a two-dimensional section of the s-velocity ball D_s. The upper circle is the image of the lower circle under the transformation $w \to \frac{a+w}{1+\bar{a}w}$. Each circle is enhanced with a grid to highlight the effect of this transformation. Notice how a typical square of the lower grid is deformed and changes in size under the transformation.

For each s-velocity $\mathbf{a} \in D_s$, we define the left translation $L_{\mathbf{a}} : D_s \to D_s$ by

$$L_{\mathbf{a}}(\mathbf{w}) = \mathbf{a} \oplus_s \mathbf{w}. \tag{29}$$

Each left translation $L_{\mathbf{a}}$ is a *conformal* automorphism of D_s. In addition, the inverse of a left translation is again a left translation. In fact $L_{\mathbf{a}}^{-1} = L_{-\mathbf{a}}$. As a result, the same argument as that in Section 3 shows that the group $Aut_c(D_s)$ of all conformal automorphisms of D_s has the following characterization:

$$Aut_c(D_s) = \{L_{\mathbf{a}}U : \mathbf{a} \in D_s,\ U \in O(3)\}. \tag{30}$$

We write $L_{\mathbf{a},U}$ instead of $L_{\mathbf{a}}U$.

The group $Aut_c(D_s)$ is a representation of the Lorentz group by conformal maps. It is a real Lie group of dimension 6, since any element of the group is determined by an element $\mathbf{a}$ of the three-dimensional open unit ball in $\mathbb{R}^3$ and an element U of the three-dimensional orthogonal group $O(3)$.

The two groups $Aut_p(D_s)$ and $Aut_c(D_s)$ are isomorphic. In fact, the isomorphism is given by

$$L_{\mathbf{v},U} \longleftrightarrow \Psi L_{\mathbf{v},U} \Psi^{-1}. \tag{31}$$

Nevertheless, they lead to different dynamics, as we will see.

The relativistic dynamics equation for symmetric velocities will be constructed from the elements of the Lie algebra $aut_c(D_s)$ of $Aut_c(D_s)$. To define the elements of $aut_c(D_s)$, consider differentiable curves $g(s)$ from a neighborhood I_0 of zero into $Aut_c(D_s)$, with $g(0) = L_{\mathbf{0},I}$, the identity of $Aut_c(D_s)$. Then

$$g(s) = L_{\mathbf{a}(s),U(s)}, \tag{32}$$

where $\mathbf{a} : I_0 \to D_s$ is a differentiable function satisfying $\mathbf{a}(0) = \mathbf{0}$ and $U(s) : I_0 \to O(3)$ is differentiable and satisfies $U(0) = I$. We denote by ξ the element of $aut_c(D_s)$ generated by $g(s)$. For any fixed $\mathbf{w} \in D_s$, $g(s)(\mathbf{w})$ is a smooth curve in D_s, with $g(0)(\mathbf{w}) = \mathbf{w}$, and $\xi(\mathbf{w})$ is a tangent vector to this line. Thus, the elements of $aut_c(D_s)$ are vector fields $\xi(\mathbf{w})$ on D_s defined by

$$\xi(\mathbf{w}) = \frac{d}{ds} g(s)(\mathbf{w})\Big|_{s=0}. \tag{33}$$

The explicit form of $\xi(\mathbf{w})$ is calculated in [1]. There it is shown that

$$aut_c(D_s) = \{\xi_{\mathbf{b},A} : \mathbf{b} \in \mathbb{R}^3, A \text{ is a } 3 \times 3 \text{ antisymmetric matrix}\}, \tag{34}$$

where

$$\xi_{\mathbf{b},A}(\mathbf{w}) = \mathbf{b} + A\mathbf{w} - 2\langle \mathbf{b}|\mathbf{w}\rangle \mathbf{w} + |\mathbf{w}|^2 \mathbf{b}. \tag{35}$$

It is useful to express the generator $\xi_{\mathbf{b},A}$ in terms of the triple product

$$\{\mathbf{a}, \mathbf{b}, \mathbf{c}\} = \langle \mathbf{a}|\mathbf{b}\rangle \mathbf{c} + \langle \mathbf{c}|\mathbf{b}\rangle \mathbf{a} - \langle \mathbf{a}|\mathbf{c}\rangle \mathbf{b}, \tag{36}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$. This product is called the *spin triple product*. The bounded symmetric domain D_s endowed with the spin triple product is called the *spin factor* and is a domain of type IV in Cartan's classification. See Chapters 2 and 3 of [1] for a full treatment of the spin triple product in the theory of Bounded Symmetric Domains and Special Relativity.

Rewriting the generators (35) using the triple product, we find that

$$aut_c(D_s) = \{\xi_{\mathbf{b},\mathbf{B}} : \mathbf{b}, \mathbf{B} \in \mathbb{R}^3\}, \tag{37}$$

where $\xi_{\mathbf{b},\mathbf{B}} : D_s \to \mathbb{R}^3$ is the vector field defined by

$$\xi_{\mathbf{b},\mathbf{B}}(\mathbf{w}) = \mathbf{b} + \mathbf{w} \times \mathbf{B} - \{\mathbf{w}, \mathbf{b}, \mathbf{w}\}. \tag{38}$$

See Figures 8 and 9 for two examples of these vector fields.

The similarities between Figures 3 and 8 and between Figures 4 and 9 can be misleading. For example, the flow generating Figure 4 is elliptical, while the trajectories in Figure 9 are circles.

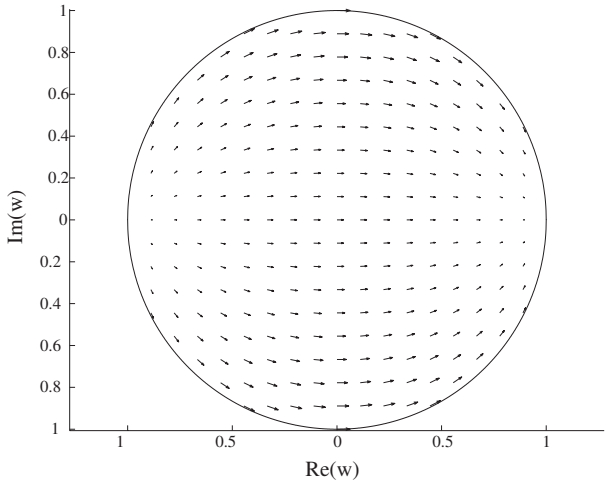

Figure 8. *The vector field of the electric field E on D_s.*
The vector field $\xi_{\mathbf{b},\mathbf{B}}$, with $\mathbf{b} = (0.07, 0, 0)$ and $\mathbf{B} = 0$, on a two-dimensional section of the s-velocity ball D_s. Note that this vector field is similar to the corresponding one for the Lie algebra $aut_p(D_v)$ of the velocity ball (see Figure 3).

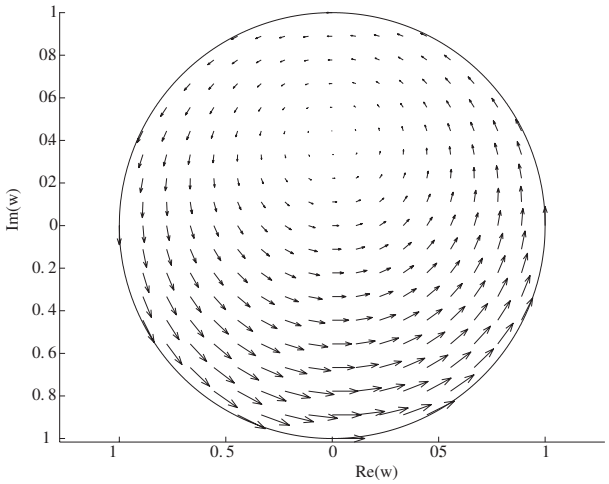

Figure 9. The vector field of the electromagnetic field $E, \mathbf{B}$ on D_s.
The vector field $\xi_{\mathbf{b},\mathbf{B}}$ with $\mathbf{b} = (0.07, 0, 0)$ and $\mathbf{B} = (0, 0, 0.1)$, on a two-dimensional section of the s-velocity ball D_s. Note that this vector field is similar to the corresponding one for the Lie algebra $aut_p(D_v)$ of the velocity ball (see Figure 4).

To obtain the relativistic dynamics equation for symmetric velocities, we start with

$$m_0 \frac{d(\gamma \mathbf{v})}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{39}$$

and change variables from velocity $\mathbf{v}$ to s-velocity $\mathbf{w}$. We obtain

$$m_0 c \frac{d\mathbf{w}(\tau)}{d\tau} = q(\mathbf{E}/2 + \mathbf{w}(\tau) \times c\mathbf{B} - \{\mathbf{w}(\tau), \mathbf{E}/2, \mathbf{w}(\tau)\}) = q\xi_{\mathbf{E}/2, c\mathbf{B}}(\mathbf{w}(\tau)). \tag{40}$$

Thus the left translations yield a nice development of relativistic dynamics also in the case of symmetric velocity. The right translations, on the other hand, are again problematic. Even in the ostensibly simpler two-dimensional version of symmetric velocity addition, the inverses of right translations are rather unwieldy. Recall that the two-dimensional version of s-velocity addition is

$$a \oplus_s b = \frac{a + b}{1 + \overline{a}b}. \tag{41}$$

The left inverses are well-behaved, and we have

$$L_a^{-1}(b) = \frac{-a + b}{1 - \overline{a}b} = L_{-a}(b). \tag{42}$$

Compare this to the right inverse, which is not even analytic:

$$R_a^{-1}(b) = \frac{b(1 - |a|^2) - a(1 - |b|^2)}{1 - |b|^2|a|^2}. \tag{43}$$

Again the right translations have lead to an apparent dead end.

6. Why are the left and right translations so different?

Why are the *left* translations so well-suited for relativistic dynamics, while the *right* translations are not? Who told forces that they have to act on the left?

Let's take another look at the physical definition of the Einstein velocity addition. Suppose an observer is at rest in an inertial system K. For any velocity $\mathbf{a} \in D_v$, let $K_{\mathbf{a}}$ denote the inertial system whose axes are parallel to those of K and moves with relative velocity $\mathbf{a}$ with respect to K. Then the inverse functions $L_{\mathbf{a}}^{-1}$ and $R_{\mathbf{a}}^{-1}$ now have the following physical

interpretation. The question "What is the value of $L_{\mathbf{a}}^{-1}(\mathbf{u})$?" is equivalent to the question "Which *velocity* measured in the system $K_{\mathbf{a}}$ will be measured by our observer as $\mathbf{u}$?" whereas the question "What is the value of $R_{\mathbf{a}}^{-1}(\mathbf{u})$?" is equivalent to the question "In which *system* K' will the velocity $\mathbf{a}$ be measured by our observer as $\mathbf{u}$?" In other words, the two inverse functions are answering fundamentally different questions.

This might explain why the left and right translations behave *differently*. But it still does not explain the preferred status of left over right.

7. Suggestions for further research

As mentioned previously, we believe that the triple product might be helpful in overcoming the difficulties inherent in using the right translations. In the two-dimensional version of s-velocity addition, for example, the triple product is derived from (36) and takes the form

$$\{z, b, w\} = z\bar{b}w, \tag{44}$$

where $z, b, w \in \mathbb{C}$. Then, although neither the right translation R_a nor its inverse is analytic, each of these functions does have a power series expansion. More explicitly, for the right translation we have

$$R_a(b) = \frac{b+a}{1+\bar{b}a} = \sum_{n=0}^{\infty}(-1)^n D(a,b)^n(a+b), \tag{45}$$

where $D(a,b)x = \{a, b, x\}$ and $D(a,b)^0 = \mathrm{Id}$. For the inverse, we have

$$R_a^{-1}(b) = \frac{b(1-|a|^2) - a(1-|b|^2)}{1-|b|^2|a|^2} = \sum_{n=0}^{\infty} Q(a,b)^n(-a+b), \tag{46}$$

where $Q(a,b)x = \{a, x, b\}$ and $Q(a,b)^0 = \mathrm{Id}$. It will be interesting to see if these power series make the right translations more amenable to relativistic dynamics.

Another line of investigation involves the *Equivalence Principle*. This principle has several versions. The classical version ([3], p. 244) states that motion in a uniformly accelerated system is the same as that in an inertial system in the presence of a gravitational field. According to the *generalized Equivalence Principle*, motion in a uniformly accelerated system is the same as that in an inertial system in the presence of *any* force. We believe that

using the left translations of either the Einstein or the symmetric velocity loop is equivalent to the generalized Equivalence Principle. In other words, the right translations are the wrong ones to use because they contradict the generalized Equivalence Principle. Moreover, if our belief is correct, then, once we succeed in developing relativistic dynamics from the right translations, we will have a way of testing the correctness of the Equivalence Principle itself. These ideas will be taken up in a forthcoming paper.

Acknowledgments. The author would like to thank Professor Yaakov Friedman for his continued guidance and the referee for his helpful remarks.

References

[1] **Y. Friedman**: *Physical Applications of Homogeneous Balls*, Progress in Mathematical Physics **40** Birkhäuser, Boston, 2004.

[2] **Y. Friedman, M. Semon**: *Relativistic acceleration of charged particles in uniform and mutually perpendicular electric and magnetic fields as viewed in the laboratory frame*, Phys. Rev. E **72** (2005), 026603.

[3] **L. Landau, E. Lifshitz**: *The Classical Theory of Fields*, Addison-Wesley, Reading, 1959.

[4] **W. Rindler**: *Relativity: Special, General, and Cosmological*, Oxford University Press, New York, 2001.

[5] **S. Takeuchi**: *Relativistic $E \times B$ acceleration*, Phys. Rev. E **66** (2002), 37402.

[6] **A. A. Ungar**: *Beyond the Einstein Addition Law and its Gyroscopic Thomas Precession*, Fundamental Theories of Physics **117**, Kluwer Academic Publisher, 2001.

Jerusalem College of Technology
Department of Mathematics
Jerusalem 91160
Israel
e-mail: scarr@jct.ac.il

Received October 26, 2006

Quasigroups and Related Systems **14** (2006), 111 – 122

Mathematical concepts of evolution algebras in non-Mendelian genetics

Jianjun Paul Tian and Petr Vojtěchovský

Abstract

Evolution algebras are not necessarily associative algebras satisfying $e_ie_j = 0$ whenever e_i, e_j are two distinct basis elements. They mimic the self-reproduction of alleles in non-Mendelian genetics. We present elementary mathematical properties of evolution algebras that are of importance from the biological point of view.

Several models of Mendelian [2, 4, 12, 6, 8, 11] and non-Mendelian genetics [1, 5] exist. Based on the self-reproduction rule of non-Mendelian genetics [1, 7], the first author introduced a new type of algebra [10], called *evolution algebra*. In this paper we discuss some basic properties of evolution algebras.

1. Evolution algebras and subalgebras

Let K be a field. A vector space E over K equipped with multiplication is an *algebra* (not necessarily associative) if $u(v+w) = uv + uw$, $(u+v)w = uw + vw$, $(\alpha u)v = \alpha(uv) = u(\alpha v)$ for every u, v, $w \in E$ and $\alpha \in K$.

Let $\{e_i;\ i \in I\}$ be a basis of an algebra E. Then $e_ie_j = \sum_{k\in I} a_{ijk}e_k$ for some $a_{ijk} \in K$, where only finitely many *structure constants* a_{ijk} are nonzero for a fixed i, $j \in I$. The multiplication in E is fully determined by the structure constants a_{ijk}, thanks to the distributive laws.

2000 Mathematics Subject Classification: 17A20, 92C15, 20N05, 37N25
Keywords: Nonassociative algebra, evolution algebra, non-Mendelian genetics, allele, self-reproduction
J. P. Tian would like to acknowledge the support from the National Science Foundation upon agreement No. 0112050

Let E be an algebra. Then $F \subseteq E$ is a *subalgebra* of E if F is a subspace of E closed under multiplication.

It is not difficult to show that the intersection of subalgebras is a subalgebra. Thus, given a subset S of E, there is the smallest subalgebra of E containing S. We call it the *subalgebra generated by* S, and denote it by $\langle S \rangle$. As usual:

Lemma 1.1. *Let S be a subset of an algebra E. Then $\langle S \rangle$ consists of all elements of the form $\alpha_1(s_{1,1} \cdots s_{1,m_1}) + \cdots + \alpha_k(s_{k,1} \cdots s_{k,m_k})$, where $k \geqslant 1$, $m_i \geqslant 0$, $s_{i,j} \in S$, $\alpha_i \in K$, and where the product $s_{i,1} \cdots s_{i,m_i}$ is parenthesized in some way.*

An *ideal* I of an algebra E is a subalgebra of E satisfying $I \cdot E \subseteq I$, $E \cdot I \subseteq I$. Clearly, 0 and E are ideals of E, called *improper* ideals. All other ideals are *proper*. An algebra is *simple* if it has no proper ideals.

An *evolution algebra* is a finite-dimensional algebra E over K with basis $\{e_1, \ldots, e_v\}$ such that $a_{ijk} = 0$ whenever $i \neq j$. Upon renaming the structure constants we can write $e_i e_i = \sum_{j=1}^{v} a_{ij} e_j$. We refer to $\{e_1, \ldots, e_v\}$ as the *natural basis* of an algebra E. An evolution algebra is *nondegenerate* if $e_i e_i \neq 0$ for every i. *Throughout the paper we will assume that evolution algebras are nondegenerate.*

The multiplication in an evolution algebra is supposed to mimic self-reproduction of non-Mendelian genetics. We think and speak of the generators e_i as *alleles*. The rule $e_i e_j = 0$ for $i \neq j$ is then natural, and the rule $e_i e_i = \sum a_{ij} e_j$ can be interpreted as follows: a_{ij} is the probability that e_i becomes e_j in the next generation, and thus $\sum a_{ij} e_j$ is the superposition of the possible states. Nevertheless, we will develop much of the theory over arbitrary fields and with no (probabilistic) restrictions on the structure constants a_{ij}.

Given two elements

$$x = \sum_{i=1}^{v} \alpha_i e_i, \qquad y = \sum_{i=1}^{v} \beta_i e_i,$$

of an evolution algebra, we have

$$\begin{aligned} xy &= \sum_{i=1}^{v} \alpha_i e_i \cdot \sum_{j=1}^{v} \beta_j e_j = \sum_{i=1}^{v} \alpha_i \beta_i e_i^2 \\ &= \sum_{i=1}^{v} \left(\alpha_i \beta_i \sum_{j=1}^{v} a_{ij} e_j \right) = \sum_{j=1}^{v} \left(\sum_{i=1}^{v} \alpha_i \beta_i a_{ij} \right) e_j, \end{aligned}$$

a formula we will use without reference.

The natural basis of an evolution algebra plays a privileged role among all other bases, since the generators e_i represent alleles. Importantly, the natural basis is privileged for purely mathematical reasons, too. The following example illustrates this point:

Example 1.2. Let E be an evolution algebra with basis $\{e_1, e_2, e_3\}$ and multiplication defined by $e_1e_1 = e_1 + e_2$, $e_2e_2 = -e_1 - e_2$, $e_3e_3 = -e_2 + e_3$. Let $u_1 = e_1 + e_2$, $u_2 = e_1 + e_3$. Then $(\alpha u_1 + \beta u_2)(\gamma u_1 + \delta u_2) = \alpha\gamma u_1^2 + (\alpha\delta + \beta\gamma)u_1u_2 + \beta\delta u_2^2 = (\alpha\delta + \beta\gamma)u_1 + \beta\delta u_2$. Hence $F = Ku_1 + Ku_2$ is a subalgebra of E. However, F is not an evolution algebra:

Let $\{v_1, v_2\}$ be a basis of F. Then $v_1 = \alpha u_1 + \beta u_2$, $v_2 = \gamma u_1 + \delta u_2$ for some $\alpha, \beta, \gamma, \delta \in K$ such that $D = \alpha\delta - \beta\gamma \neq 0$. By the above calculation, $v_1v_2 = (\alpha\delta + \beta\gamma)u_1 + \beta\delta u_2$. Assume that $v_1v_2 = 0$. Then $\beta\delta = 0$ and $\alpha\delta + \beta\gamma = 0$. If $\beta = 0$, we have $\alpha\delta = 0$. But then $D = 0$, a contradiction. If $\delta = 0$, we reach the same contradiction. Hence $v_1v_2 \neq 0$, and F is not an evolution algebra.

We have just seen that evolution algebras are not closed under subalgebras. We therefore say that a subalgebra F of an evolution algebra E with basis $\{e_1, \ldots, e_v\}$ is an *evolution subalgebra* if, as a vector space, it is spanned by $\{e_i;\ i \in I\}$ for some $I \subseteq \{1, \ldots, v\}$. The subset I determines F uniquely, and we write $F = E(I) = \{\sum_{i=1}^{v} \alpha_i e_i;\ \alpha_i = 0 \text{ when } i \notin I\}$.

Similarly, we define an *evolution ideal* as an ideal I of E that happens to be an evolution subalgebra. This concept is superfluous, however:

Lemma 1.3. *Every evolution subalgebra is an evolution ideal.*

Proof. Let $F = E(I)$ be an evolution subalgebra . Let $x = \sum_{i\in I} \alpha_i e_i$ be an element of F and e_j an allele. We need to show that $xe_j \in F$. When $j \notin I$ then $xe_j = 0 \in F$. Assume that $j \in I$. Since F is an evolution subalgebra, $e_i \in F$ for every $i \in I$. Then $xe_j = \alpha_j e_j^2 \in F$, since F is a subalgebra. □

Not every ideal of an evolution algebra is an evolution ideal:

Example 1.4. *Let E be generated by e_1, e_2, where $e_1e_1 = e_1 + e_2 = e_2e_2$. Then $K(e_1 + e_2)$ is an ideal of E, but not an evolution subalgebra.*

An evolution algebra is *evolutionary simple* if if has no proper evolution ideals (evolution subalgebras).

Clearly, every simple evolution algebra is evolutionary simple. The converse it not true, as is apparent from Example 1.4.

The following theorem gives some basic properties of evolution algebras, all easy to prove (or see [10]). Recall that an algebra is *flexible* if it satisfies $x(yx) = (xy)x$.

Theorem 1.5. *Evolution algebras are commutative (and hence flexible), but not necessarily power-associative (hence not necessarily associative). Direct products and direct sums of evolution algebras are evolution algebras. Evolution subalgebra of an evolution algebra is an evolution algebra.*

An algebra is *real* if $K = \mathbb{R}$. An evolution algebra is *nonnegative* if it is real and all structure constants a_{ij} are nonnegative. A *Markov evolution algebra* is a nonnegative evolution algebra such that $\sum_j a_{ij} = 1$ for every $1 \leqslant i \leqslant v$.

When E is a real algebra, let $E^+ = \{\sum \alpha_i e_i;\ \alpha_i \geqslant 0\}$.

Lemma 1.6. *Let E be a nonnegative evolution algebra. Then E^+ is closed under addition, multiplication, and multiplication by positive scalars.*

Proof. Let $x = \sum \alpha_i e_i$, $y = \sum \beta_i e_i \in E^+$. Then $x + y = \sum(\alpha_i + \beta_i)e_i$ clearly belongs to E^+. Moreover, $xy = \sum_j \left(\sum_i \alpha_i\beta_i a_{ij}\right) e_j \in E^+$, since α_i, β_i, $a_{ij} \geqslant 0$ for every i, j. It is clear that E^+ is closed under multiplication by nonnegative scalars. □

2. The evolution operator

Let E be an evolution algebra with basis $\{e_1, \ldots, e_v\}$. Since we are mainly interested in self-reproduction, we focus on the *evolution operator* $\Lambda : E \to E$, which is the (unique) linear extension of the map $e_i \mapsto e_i^2$.

Lemma 2.1. *Let E be an evolution algebra and $x = \sum \alpha_i e_i$. Then $\Lambda(x) = x^2$, i.e., $\sum \alpha_i^2 e_i^2 = (\sum \alpha_i e_i)^2$.*

Proof. This is an immediate consequence of the fact that $e_i e_j = 0$ whenever $i \neq j$. □

When E is a real evolution algebra, we can equip it with the usual L_1 norm, i.e., $\|\sum \alpha_i e_i\| = \sum |\alpha_i|$. Since E is then isomorphic to $\mathbb{R}^v$ as a vector space, it becomes a complete vector space with respect to the metric $d(x, y) = \|x - y\|$. In other words, E is a Banach space.

Since $v < \infty$, all linear operators defined on E are continuous. In particular, every *left translation by z*, defined by $L_z(x) = zx$, is a continuous

operator on E. However, due to the lack of associativity, the composition of two left translations does not have to be a left translation.

A (not-necessarily associative) *Banach algebra* is an algebra that is also a Banach space with norm $\|\cdot\|$ satisfying $\|xy\| \leqslant \|x\| \cdot \|y\|$. Not every evolution algebra is a Banach algebra. However:

Lemma 2.2. *Let E be a real evolution algebra such that $\sum_j |a_{ij}| \leqslant 1$ for every i (eg. a Markov evolution algebra). Then E is a Banach algebra.*

Proof. Let $x = \sum_i \alpha_i e_i$, $y = \sum_i \beta_i e_i$. Then $\|x\| \cdot \|y\| = \sum_i |\alpha_i| \cdot \sum_j |\beta_j|$. On the other hand,

$$\|xy\| = \left\| \sum_j \left(\sum_i \alpha_i \beta_i a_{ij} \right) e_j \right\| = \sum_j \left| \sum_i \alpha_i \beta_i a_{ij} \right| \leqslant \sum_j \sum_i \left(|\alpha_i \beta_i| \cdot |a_{ij}| \right)$$
$$= \sum_i \left(\sum_j |a_{ij}| \right) |\alpha_i \beta_i| \leqslant \sum_i |\alpha_i \beta_i|,$$

and the needed inequality follows. □

Note that even in the case of a Markov evolution algebra we never have $\|xy\| = \|x\| \cdot \|y\|$ for every x, y, as long as $v > 1$. For instance, $\|e_i e_j\| = 0 < 1 = \|e_i\| \cdot \|e_j\|$ when $i \neq j$.

Given x in an algebra E, we define the *plenary powers* of x by $x^{[0]} = x$, $x^{[n+1]} = x^{[n]} x^{[n]}$. Equivalently, we can set $x^{[n]}$ equal to $\Lambda^n(x)$ for any $n \geqslant 0$.

Recall that composition of maps is an associative binary operation. Thus:

Lemma 2.3. *Let E be an algebra, $x \in E$, $\alpha \in K$, and n, $m \geqslant 0$. Then:*

(*i*) $(x^{[n]})^{[m]} = x^{[n+m]}$,

(*ii*) $(\alpha x)^{[n]} = \alpha^{(2^n)} x^{[n]}$.

Proof. It remains to prove (ii), which is easy by an induction on n. □

3. Occurrence relation

The question we are most interested in is the following: *When does the allele e_i give rise to the allele e_j?* The phrase *give rise* can be interpreted in two

ways: (i) the self-reproduction of e_i yields e_j with nonzero probability after a given number of generations, or (ii) the self-reproduction of e_i yields e_j with nonzero probability after some number of generations.

The first interpretation is studied below, while the second interpretation is investigated later, starting with Section 5.

Let E be an algebra with basis $\{e_1, \ldots, e_v\}$. We say that e_i *occurs* in $x \in E$ if the coefficient $\alpha_i \in K$ is nonzero in $x = \sum_{j=1}^{v} \alpha_j e_j$. When e_i occurs in x we write $e_i \prec x$.

Lemma 3.1. *Let E be a nonnegative evolution algebra. Then for every x, $y \in E^+$ and $n \geqslant 0$ there is $z \in E^+$ such that $(x+y)^{[n]} = x^{[n]} + z$.*

Proof. We proceed by induction on n. We have $(x+y)^{[0]} = x+y = x^{[0]}+y$, and it suffices to set $z = y$. Also, $(x+y)^{[1]} = (x+y)(x+y) = x^{[1]}+2xy+y^2$. By Lemma 1.6, $2xy + y^2 = z$ belongs to E^+.

Assume the claim is true for some $n \geqslant 1$. In particular, given $x, y \in E^+$, let $w \in E^+$ be such that $(x+y)^{[n]} = x^{[n]} + w$. Then $(x+y)^{[n+1]} = ((x+y)^{[n]})^{[1]} = (x^{[n]} + w)^{[1]}$. Since $w \in E^+$ and $x^{[n]} \in E^+$ by Lemma 1.6, we have $(x^{[n]} + w)^{[1]} = (x^{[n]})^{[1]} + z = x^{[n+1]} + z$ for some $z \in E^+$. □

Proposition 3.2. *Let E be a nonnegative evolution algebra. When $e_i \prec e_j^{[n]}$ and $e_j \prec e_k^{[m]}$ then $e_i \prec e_k^{[n+m]}$.*

Proof. We have $e_k^{[m]} = \alpha_j e_j + y$ for some $\alpha_j \neq 0$ and $y \in E$ such that $e_j \not\prec y$. Moreover, by Lemma 1.6, we have $\alpha_j > 0$ and $y \in E^+$. By Lemma 3.1, $e_k^{[n+m]} = (e_k^{[m]})^{[n]} = (\alpha_j e_j + y)^{[n]} = (\alpha_j e_j)^{[n]} + z = \alpha_j^{(2^n)} e_j^{[n]} + z$ for some $z \in E^+$. Now, $e_j^{[n]} = \beta_i e_i + v$ for some $\beta_i > 0$ and $v \in E$ satisfying $e_i \not\prec v$. We therefore conclude that $e_i \prec e_k^{[n+m]}$. □

The proposition does not generalize to all evolution algebras, as the following example shows:

Example 3.3. Let E be an evolution algebra with basis $\{e_i;\ 1 \leqslant i \leqslant 7\}$ such that $e_1e_1 = e_1$, $e_2e_2 = e_4$, $e_3e_3 = e_5 + e_6$, $e_4e_4 = e_1$, $e_5e_5 = e_2$, $e_6e_6 = e_7$, $e_7e_7 = -e_4$. Then $e_2^{[1]} = e_2e_2 = e_4$, $e_2^{[2]} = e_4e_4 = e_1$. Thus $e_1 \prec e_2^{[2]}$. Also, $e_3^{[1]} = e_3e_3 = e_5 + e_6$, $e_3^{[2]} = (e_5 + e_6)^2 = e_5^2 + e_6^2 = e_2 + e_7$. Thus $e_2 \prec e_3^{[2]}$. However, $e_3^{[3]} = (e_2 + e_7)^2 = e_2^2 + e_7^2 = e_4 - e_4 = 0$, and so $e_3^{[n]} = 0$ for every $n \geqslant 3$. This means that $e_1 \not\prec e_3^{[n]}$ for any $n \geqslant 0$.

4. Occurrence sets

Let e_i, e_j be two alleles of an evolution algebra. Then the *occurrence set of* e_i *with respect to* e_j is the set $O_{i,j} = \{n > 0;\ e_i \prec e_j^{[n]}\}$.

Recall that a *semigroup* is a set with one binary operation that satisfies the associative law. When E is a nonnegative evolution algebra, every occurrence set $O_{i,i}$ is a subsemigroup of $(\{1, 2, \dots\}, +)$ by Proposition 3.2.

The goal of this section is to show that any finite subset of $\{1, 2, \dots\}$ can be realized as an occurrence set of some evolution algebra, and that every subsemigroup of $(\{1, 2, \dots\}, +)$ can be realized as an occurrence set of some nonnegative evolution algebra. Hence the occurrence sets are as rich as one could hope for.

Example 4.1. *Let* $n > 1$. *Consider the evolution algebra* E *with generators* $\{e_1, \dots, e_{n+1}\}$ *defined by* $e_1e_1 = e_2$, $e_2e_2 = e_3$, $\dots$, $e_{n-1}e_{n-1} = e_n$, $e_ne_n = e_1 + e_{n+1}$, $e_{n+1}e_{n+1} = -e_2$. *Then* $e_1^{[m]} = e_{m+1}$ *for every* $1 \leqslant m < n$, $e_1^{[n]} = e_1 + e_{n+1}$, *and* $e_1^{[m]} = 0$ *for every* $m > n$. *Thus* $O_{1,1} = \{n\}$.

Lemma 4.2. *Let* S *be a finite subset of* $\{1, 2, \dots\}$. *Then there is an evolution algebra* E *such that* $O_{1,1} = S$.

Proof. Let $S = \{n_1, \dots, n_m\}$. In the following calculations we label basis elements of E also by $e_{i,j}$; these can be relabeled as e_i at the end.

Let $e_1e_1 = e_{2,1} + \cdots + e_{2,m}$. Given $1 \leqslant i \leqslant m$, let $e_{2,i}e_{2,i} = e_{3,i}$, $e_{3,i}e_{3,i} = e_{4,i}$, $\dots$, $e_{n_i,i}e_{n_i,i} = e_1 + e_{n_i+1,i}$, $e_{n_i+1,i}e_{n_i+1,i} = -e_1e_1$. Thus, roughly speaking, we imitate Example 4.1 for every $1 \leqslant i \leqslant m$. It is now not hard to see that $O_{1,1} = S$. □

A semigroup is *finitely generated* if it is generated by a finite subset. Here is a well-known fact:

Lemma 4.3. *Every subsemigroup of* $(\{1, 2, \dots\}, +)$ *is finitely generated.*

Proof. Let S be a subsemigroup of $(\{1, 2, \dots\}, +)$. Let n be the smallest element of S. For every $1 \leqslant i < n$ let m_i be the smallest element of S such that m_i is congruent to i modulo n, if such an element exists, else set $m_i = n$. We claim that $A = \{n, m_1, \dots, m_{n-1}\}$ generates S. Suppose that this is not the case and let s be the smallest element of S not generated by A. Since s cannot be a multiple of n, there is $1 \leqslant i < n$ such that s is congruent to i modulo n. Then $m_i \neq n$ and $m_i < s$. But then $s = m_i + kn$ for some $k > 0$, so $s \in A$, a contradiction. □

Lemma 4.4. *Let S be a subsemigroup of $(\{1, 2, \dots\}, +)$. Then there is a nonnegative evolution algebra E such that $O_{1,1} = S$.*

Proof. Assume that S is 1-generated, i.e., that $S = \{n, 2n, \dots\}$ for some $n \geqslant 1$. Then define E by: $e_1e_1 = e_2$, $e_2e_2 = e_3$, $\dots$, $e_{n-1}e_{n-1} = e_n$, $e_ne_n = e_1$. It is easy to see that $O_{1,1} = S$.

When S is generated by m elements, say $n_1, \dots, n_m$, we can use a similar trick as in the proof of Lemma 4.2.

Every subsemigroup of $(\{1, 2, \dots\}, +)$ is finitely generated by Lemma 4.3. □

Problem 4.5. *Can any subset of $\{1, 2, \dots\}$ be realized as an occurrence set of some evolution algebra?*

Problem 4.6. *Let S be a subset of $\{1, 2, \dots\}$, $|S| = n$. What is the smallest integer v such that there is an evolution algebra E of dimension v for which S is an occurrence set?*

5. Occurrence based on evolution subalgebras

We are now going to look at the second interpretation of "e_i gives rise to e_j."

Lemma 5.1. *Intersection of evolution subalgebras is an evolution subalgebra.*

Proof. Let $F = E(I)$, $G = E(J)$ be two evolution subalgebras of E. Then $F \cap G = E(I \cap J)$ as a vector space. Since $F \cap G$ is a subalgebra, we are done. □

Thus for any subset S of E there exists the smallest evolution subalgebra of E containing S, and we denote it by $\langle\langle S\rangle\rangle$. The notation is supposed to suggest that the evolution subalgebra generated by S can be larger than the subalgebra generated by S.

We now define another occurrence relation as follows: For $x, y \in E$, let $x \ll y$ if $x \in \langle\langle y\rangle\rangle$.

Lemma 5.2. *For $x, y, z \in E$ we have:*

(*i*) *if $x \ll y$ and $y \ll x$ then $\langle\langle x\rangle\rangle = \langle\langle y\rangle\rangle$,*

(*ii*) *if $x \ll y$ and $y \ll z$ then $x \ll z$,*

(*iii*) *if* $x \ll y^{[n]}$ *for some* $n \geqslant 0$ *then* $x \ll y$.

Proof. Easy. □

In view of Lemma 5.2(*iii*), it makes no sense to speak of occurrence sets (analogous to $O_{i,j}$) in the context of $\ll$, since every occurrence set would be either empty or would consists of all nonnegative integers.

Lemma 5.3. *Let* F, G *be evolutionary simple evolution subalgebras of* E. *Then either* $F = G$ *or* $F \cap G = 0$.

Proof. Assume that there is $x \in F \cap G$, $x \neq 0$. Then $\langle\langle x \rangle\rangle$ is an evolution subalgebra of both F and G. Since both F, G are evolutionary simple, it follows that $F = G = \langle\langle x \rangle\rangle$. □

6. Algebraically persistent and transient generators

A generator e_i of an evolution algebra E is *algebraically persistent* if $\langle\langle e_i \rangle\rangle$ is evolutionary simple, else it is *algebraically transient*.

Lemma 6.1. *If* E *is an evolutionary simple evolution algebra then it has no algebraically transient generators.*

Proof. Assume that e_i is an algebraically transient generator, i.e., that $\langle\langle e_i \rangle\rangle$ is not evolutionary simple. If $E = \langle\langle e_i \rangle\rangle$, we see right away that E is not evolutionary simple. If $\langle\langle e_i \rangle\rangle$ is a proper evolution subalgebra of E then it is a proper evolution ideal of E by Lemma 1.3, and E is not evolutionary simple. □

The following example shows that the converse of Lemma 6.1 does not hold (but see Corollary 7.3):

Example 6.2. Let E have generators e_1, e_2 such that $e_1e_1 = e_1$, $e_2e_2 = e_2$. Then $\langle\langle e_1 \rangle\rangle = Ke_1$, $\langle\langle e_2 \rangle\rangle = Ke_2$, which means that both e_1, e_2 are algebraically persistent. Yet $\langle\langle e_i \rangle\rangle$ is a proper evolution ideal of E, and hence E is not evolutionary simple.

Lemma 6.3. *Let* e_i *be an algebraically persistent generator of* E, *and assume that* $e_j \prec e_ie_i$. *Then* e_j *is algebraically persistent.*

Proof. Since $e_j \prec e_ie_i$, we have $\langle\langle e_i \rangle\rangle \supseteq \langle\langle e_j \rangle\rangle$. But $\langle\langle e_i \rangle\rangle$ is evolutionary simple, thus $\langle\langle e_i \rangle\rangle = \langle\langle e_j \rangle\rangle$. Then $\langle\langle e_j \rangle\rangle$ is evolutionary simple, and thus e_j is algebraically persistent. □

7. Decomposition of evolution algebras

An evolution algebra E is *indecomposable* if whenever $E = F \oplus G$ for some evolution subalgebras F, G of E, we have $F = 0$ or $G = 0$. An easy induction proves that every evolution algebra can be written as a direct sum of indecomposable evolution algebras.

Here is an indecomposable evolution algebra that is not evolutionary simple:

Example 7.1. Let E be generated by e_1, e_2, where $e_1e_1 = e_1$, $e_2e_2 = e_1$. Then $\langle\langle e_1\rangle\rangle = Ke_1$, $\langle\langle e_2\rangle\rangle = E$.

An evolution algebra E is *evolutionary semisimple* if it is a direct sum of some of its evolutionary simple evolution subalgebras. Note that every evolutionary simple evolution subalgebra of E can be written as $\langle\langle e_i\rangle\rangle$ for some algebraically persistent generator of E.

Proposition 7.2. *An evolution algebra E is evolutionary semisimple if and only if all of its alleles e_i are algebraically persistent.*

Proof. Assume that E is evolutionary semisimple, and write $E = \langle\langle e_{i_1}\rangle\rangle \oplus \cdots \oplus \langle\langle e_{i_n}\rangle\rangle$, where each e_{i_j} is algebraically persistent. Let e_j be an allele of E. Then e_j belongs to some $\langle\langle e_{i_k}\rangle\rangle$. Since $\langle\langle e_j\rangle\rangle$ is an evolution ideal of $\langle\langle e_{i_k}\rangle\rangle$ and e_{i_k} is algebraically persistent, we conclude that $\langle\langle e_j\rangle\rangle = \langle\langle e_{i_k}\rangle\rangle$. Thus e_j is algebraically persistent, too.

Conversely, assume that every allele of E is algebraically persistent. For each e_i let $I_i = \{j;\ e_j \ll e_i\}$. Given $i \neq j$, we have either $I_i = I_j$ or $I_i \cap I_j = \emptyset$, by Lemma 5.3. Thus there exists $\{i_1, \ldots, i_n\} \subseteq \{1, \ldots, v\} = I$ such that $I_{i_1} \cup \cdots \cup I_{i_n} = I$, and the union is disjoint. In other words, $E = \langle\langle e_{i_1}\rangle\rangle \oplus \cdots \oplus \langle\langle e_{i_n}\rangle\rangle$. □

Here is a partial converse of Lemma 6.1:

Corollary 7.3. *An indecomposable evolution algebra with no transient generators is evolutionary simple.*

Let E be an evolution algebra. Partition $\{1, \ldots, v\}$ as $I \cup J$, where $e_i \in I$ if and only if e_i is an algebraically persistent generator of E. Let $P(E) = \{\sum \alpha_i e_i;\ \alpha_i = 0 \text{ for } i \notin I\}$, and $T(E) = \{\sum \alpha_i e_i;\ \alpha_i = 0 \text{ for } i \notin J\}$.

Lemma 7.4. *$P(E)$ is an evolutionary semisimple evolution subalgebra of E.*

Proof. We first show that $P(E)$ is an evolution subalgebra. Let $x \in P(E)$, $y \in P(E)$, $x = \sum_{i\in I} \alpha_i e_i$, $y = \sum_{i\in I} \beta_i e_i$, where I is as above. Then $xy = \sum_{i\in I} \alpha_i\beta_i e_i^2$. By Lemma 6.3, e_i^2 is a linear combination of algebraically persistent generators, and hence $xy \in P(E)$.

Then $P(E)$ is evolutionary semisimple by Proposition 7.2. □

Observe:

Lemma 7.5. *Let $E(I)$, $E(J)$ be evolution subalgebras of E such that $E(I)$ is a subalgebra of $E(J)$. Then $I \subseteq J$. If $E(I)$ is a proper subalgebra of $E(J)$, then I is a proper subset of J.*

Thus:

Lemma 7.6. *Every evolution algebra E has an evolutionary simple evolution subalgebra. In particular, $P(E) \neq 0$.*

Proof. We proceed by induction on v. If $v = 1$, then $E = \langle\langle e_1\rangle\rangle$ is evolutionary simple. Assume that the lemma is true for $v - 1$. If $E = E(\{1, \dots, v\})$ is evolutionary simple, we are done. Else, by Lemma 7.5, there is a proper subset I of $\{1, \dots, v\}$ such that $E(I)$ is a proper evolution subalgebra. By induction, $E(I)$ contains an evolutionary simple evolution subalgebra. □

Every evolution algebra E decomposes as a vector space into $P(E) \oplus T(E)$, and $P(E) \neq 0$, by the above lemma. Moreover, $P(E)$ is an evolutionary semisimple evolution algebra, and can therefore be written as a direct sum of evolutionary simple evolution algebras $\langle\langle e_{i_j}\rangle\rangle$.

However, the subspace $T(E)$ does not need to be a subalgebra of E, hence it does not need to be an evolution algebra. But we can make it into an evolution algebra:

Let $T(E) = \{\sum \alpha_i e_i;\ \alpha_i = 0 \text{ for } i \notin J\}$. Let $J^* = J \setminus \{j;\ e_j^2 \subseteq P(E)\}$. (This will guarantee that the resulting evolution algebra is nondegenerate.) Let $T^*(E)$ be defined on the subspace generated by $\{e_i;\ i \in J^*\}$ by $e_i e_i = \sum_{j\in J^*} a_{ij} e_j$, where the structure constants a_{ij} are inherited from E. If $J^* \neq \emptyset$, then $T^*(E)$ is a nondegenerate evolution algebra. If $J^* = \emptyset$ then all algebraically transient generators of E vanish after the first reproduction, and therefore have no impact, biologically speaking.

If $E_1 = T^*(E) \neq 0$, we can iterate the decomposition and form $P(E_1)$, $T(E_1)$ and $T^*(E_1)$, etc. Eventually we reach a point n when $T^*(E_n) = 0$, i.e., every transient generator of E_n disappears after the first generation.

Let us emphasize that the decomposition of E thus obtained results in an evolution algebra not necessarily isomorphic to E; some information may be lost in the decomposition $P(E) \oplus T(E)$.

References

[1] **C. W. Birky, Jr.**: *The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models*, Annu. Rev. Genet. **35** (2001), 125–148.

[2] **I. M. H. Etherington**: *Non-associative algebra and the symbolism of genetics*, Proc. Roy. Soc. Edinburgh B **61** (1941), 24–42.

[3] **R. Costa, A. Grishkov, H. Quzzo and L. Peresi (editors)**: Nonassociative algebra and its applications, the Fourth International Conference, Marcel Dekker, Inc., 2000.

[4] **P. Holgate**: *Selfing in gentic algebras*, J. Math. Biology **6** (1978), 197–206.

[5] **F. Ling and T. Shibata**: *Mhr1p-dependent concatemeric mitochondrial DNA formation for generating yeast mitochondrial homoplasmic cells*, Mol. Biol. Cell **15** (Jan. 2004), 310–322.

[6] **Y. I. Lyubich**: Mathematical structure in population genetics, Springer-Verlag, New York, 1992.

[7] **G. Mendel**: *Experiments in plant-hybridization*, Classic papers in Genetics, 1–20, J. A. Petr (editor), Prentice-Hall Inc. 1959.

[8] **M. L. Reed**: *Algebraic structure of genetic inheritance*, Bull. of Amer. Math. Soc **34** ((2)) (1997), 107–130.

[9] **R. Schafer**: An introduction to nonassociative algebra, Dover Publications, Inc., New York, 1994.

[10] **J. P. Tian**: Introduction to evolution algebras, in preparation.

[11] **J. Tian and B.-L. Li**: *Coalgebraic structure of genetic inheritance*, Mathematical Bioscience and Engineering, vol. **1**, **2** (2004).

[12] **A. Worz-Busekros**: Algebras in Genetics, Lecture Notes in Biomathematics **36**, Springer-Verlag, Berlin, 1980.

Received October 10, 2005

J. P. Tian
Mathematical Biosciences Institute
The Ohio State University
Columbus, OH 43210
U.S.A.
e-mail: tianjj@mbi.ohio-state.edu

P. Vojtěchovský
Department of Mathematics
University of Denver
Denver, CO 80208
U.S.A.
e-mail: petr@math.du.edu