

# UNIT EQUATIONS ON QUATERNIONS

YIFENG HUANG

ABSTRACT. A classical result about unit equations says that if  $\Gamma_1$  and  $\Gamma_2$  are finitely generated subgroups of  $\mathbb{C}^\times$ , then the equation  $x + y = 1$  has only finitely many solutions with  $x \in \Gamma_1$  and  $y \in \Gamma_2$ . We study a noncommutative analogue of the result, where  $\Gamma_1, \Gamma_2$  are finitely generated subsemigroups of the multiplicative group of a quaternion algebra. We prove an analogous conclusion when both semigroups are generated by algebraic quaternions with norms greater than 1 and one of the semigroups is commutative. As an application in dynamics, we prove that if  $f$  and  $g$  are endomorphisms of a curve  $C$  of genus 1 over an algebraically closed field  $k$ , and  $\deg(f), \deg(g) \geq 2$ , then  $f$  and  $g$  have a common iterate if and only if some forward orbit of  $f$  on  $C(k)$  has infinite intersection with an orbit of  $g$ .

## 1. INTRODUCTION

A classical result about unit equations states that the equation  $f + g = 1$  has only finitely many solutions in a given finitely generated semigroup  $\Gamma$  in  $K^\times$ , where  $K$  is a field of characteristic zero. Unit equations have had important applications in many areas of mathematics, including Diophantine geometry ([9, 11]), arithmetic dynamics [4, p. 291] and variants of the Mordell–Lang conjecture (for instance, see [4, p. 321]). Extensions of the classical result have also been studied, for example, see [10, 16] in the characteristic  $p$  setting.

In this paper we present a class of semigroups in the standard quaternion algebra over  $\mathbb{R}$  for which the finiteness of solutions of the unit equation holds. This is the first analogous result in the noncommutative setting. In light of the many applications of unit equations, this raises the intriguing possibility that some of those applications might have noncommutative analogues.

Let  $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$  denote the quaternion algebra  $\mathbb{H}$  over  $\mathbb{R}$ , with the standard multiplication law  $i^2 = j^2 = k^2 = -1, ij = -ji = k, jk = -kj = i, ki = -ik = j$ . For an element  $\alpha = a + bi + cj + dk \in \mathbb{H}$ , where  $a, b, c, d \in \mathbb{R}$ , define its conjugation to be  $\bar{\alpha} = a - bi - cj - dk$ , its norm to

---

We thank Jason Bell, Dragos Ghioca, Tom Tucker and Michael Zieve for helpful conversations and useful comments. We thank the referee for the time devoted to reading the paper to its details, for valuable suggestions to the paper and Conjecture 1.4, and especially for generously sketching the proof of Proposition 1.5 and Proposition A.2 and informing the author about numerous useful results. This work was done with the support of Rackham One-term Dissertation Fellowship, Indu and Gopal Prasad Family Fund, and National Science Foundation grant DMS-1601844.

be  $N(\alpha) = \alpha\bar{\alpha} = \bar{\alpha}\alpha = a^2 + b^2 + c^2 + d^2$ , and its trace  $tr(\alpha) = \alpha + \bar{\alpha} = 2a$ . Write  $|\alpha| = \sqrt{N(\alpha)}$ .

We say that a quaternion  $\alpha = a + bi + cj + dk \in \mathbb{H}$  is *algebraic* if all coordinates  $a, b, c, d$  are algebraic over  $\mathbb{Q}$ . This is equivalent to requiring that  $\alpha$  satisfies a polynomial equation with coefficients in  $\mathbb{Q}$ , or that  $\mathbb{Q}[\alpha]$  is a finite field extension of  $\mathbb{Q}$ . Indeed,  $\alpha$  always satisfies the quadratic equation

$$X^2 - tr(\alpha)X + N(\alpha) = 0$$

and if  $a, b, c, d \in \bar{\mathbb{Q}}$ , then so are  $tr(\alpha)$  and  $N(\alpha)$ .

Denote by  $\mathbb{H}_a$  the subalgebra of all quaternions that are algebraic.

**Theorem 1.1.** *Let  $\Gamma_1, \Gamma_2$  be semigroups of  $\mathbb{H}_a^\times$  generated by finitely many elements of norms greater than 1, and fix  $a, a', b, b' \in \mathbb{H}_a^\times$ . If  $\Gamma_1$  is commutative, then the equation*

$$afa' + bgb' = 1$$

*has only finitely many solutions with  $f \in \Gamma_1$  and  $g \in \Gamma_2$ . In fact, for solutions  $(f, g)$ , we have effectively computable upper bounds for  $|f|, |g|$  that depend only on  $a, a', b, b'$  and generators of  $\Gamma_1, \Gamma_2$ .*

We emphasize that even though  $\Gamma_1$  is commutative, the semigroup  $\Gamma_2$  need not be commutative, and that  $a, a'$  and  $\Gamma_1$  typically will not commute with each other. The proof relies on the following result, which implies that if a certain quaternion unit equation has infinitely many solutions, then so does another equation of a different type. We note that Theorem 1.2 applies in greater generality than Theorem 1.1, as Theorem 1.2 does not require  $\Gamma_1$  to be commutative.

**Theorem 1.2.** *Let  $\Gamma_1, \Gamma_2$  be semigroups of  $\mathbb{H}_a^\times$  generated by finitely many elements of norms greater than 1, and fix  $a, a', b, b' \in \mathbb{H}_a^\times$ . Then the equation*

$$afa' + bgb' = 1$$

*has only finitely many solutions with  $f \in \Gamma_1$  and  $g \in \Gamma_2$  such that  $|1 - afa'| \neq |afa'|$ . In fact, for such pairs  $(f, g)$ , we have effectively computable upper bounds for  $|f|, |g|$  that depend only on  $a, a', b, b'$  and generators of  $\Gamma_1, \Gamma_2$ .*

Given Theorem 1.2, in order to prove Theorem 1.1 it suffices to prove the next result which involves only the semigroup  $\Gamma_1$ :

**Theorem 1.3.** *Let  $\Gamma$  be a semigroup generated by finitely many elements in  $\mathbb{H}_a$  with norms greater than 1, and fix  $a, a' \in \mathbb{H}_a^\times$ . If  $\Gamma$  is commutative, then the equation*

$$|1 - afa'| = |afa'|$$

*has only finitely many solutions with  $f \in \Gamma$ . In fact, for solutions  $f \in \Gamma$ , we have an effectively computable upper bound for  $|f|$  that depends only on  $a, a'$  and generators of  $\Gamma$ .*

We remark that Theorem 1.3 is the only step in the proof of Theorem 1.1 that uses the commutativity of  $\Gamma_1$ , so any generalization of Theorem 1.3 would immediately yield a generalization of Theorem 1.1.

In light of the above results, we make the following conjecture about noncommutative unit equations:

**Conjecture 1.4.** Let  $\Gamma_1, \Gamma_2$  be finitely generated semigroups of the multiplicative group  $A^\times$  of a finite dimensional division algebra  $A$  over  $\mathbb{Q}$ . Then for any fixed  $a, a', b, b' \in A^\times$ , the unit equation  $afa' + bgb' = 1$  has only finitely many solutions with  $f \in \Gamma_1$  and  $g \in \Gamma_2$ .

Moreover, there is an effectively computable finite subset  $S \subseteq \Gamma_1 \times \Gamma_2$  in terms of  $a, a', b, b'$  and generators of  $\Gamma_1, \Gamma_2$ , such that all solutions  $(f, g) \in \Gamma_1 \times \Gamma_2$  must lie in  $S$ .

The referee kindly points out that the ineffective part of the conjecture is true in the case where all the semigroups are commutative:

**Proposition 1.5.** *Let  $k$  be a field of characteristic zero and  $A$  be a finite-dimensional division algebra over  $k$ . Let  $\Gamma_1, \dots, \Gamma_m$  be abelian and finitely generated subgroups of the multiplicative group  $A^\times$ . Then for any fixed  $a_1, \dots, a_m, b_1, \dots, b_m \in A^\times$ , the unit equation*

$$a_1 f_1 b_1 + \dots + a_m f_m b_m = 1$$

*has only finitely many nondegenerate solutions  $(f_1, \dots, f_m) \in \Gamma_1 \times \dots \times \Gamma_m$ , i.e., solutions such that no proper subsum equals 1.*

This, of course, implies the case where all  $\Gamma_i$  are *virtually* abelian, in the sense that  $\Gamma_i$  has a finite-index abelian subgroup. Moreover, as is mentioned by the referee, every semigroup that does not contain a free semigroup of rank two is contained in a virtually abelian subgroup of  $A^\times$  (see Proposition A.2), so Proposition 1.5 also holds if  $\Gamma_i$  does not contain a free semigroup of rank two. We emphasize that Theorem 1.1 is the only currently proven case of Conjecture 1.4 where some of the semigroups are not contained in virtually abelian subgroups of  $A^\times$ . It is also the only known case where  $A$  is noncommutative and one knows an effectively computable finite set that contains all the solutions.

In Section 6, we will discuss a possible  $p$ -adic approach to Conjecture 1.4, and will give a counterexample to the matrix algebra analogue of Conjecture 1.4 in Example 6.1.

Our main theorem has the following consequence about intersections of orbits of endomorphisms of a genus-1 curve in arbitrary characteristic.

**Corollary 1.6.** *Let  $E$  be an elliptic curve over an algebraically closed field  $k$ , and let  $f, g : E \rightarrow E$  be regular maps of degrees greater than 1. If there are points  $A, B \in E(k)$  such that the forward orbits  $O_f(A) := \{A, f(A), f^2(A), \dots\}$  and  $O_g(B) := \{B, g(B), g^2(B), \dots\}$  have infinite intersection, then  $f$  and  $g$  have a common iterate, namely,  $f^{m_0} = g^{n_0}$  for some positive integers  $m_0, n_0$ .*

In fact, if  $O_f(A) \cap O_g(B)$  is nonempty, let  $m_0, n_0$  be integers such that  $f^{m_0}(A) = g^{n_0}(B)$ . Then there is an effectively computable constant  $M$  in terms of  $A, B, f, g, m_0, n_0$  such that, if  $f^m(A) = g^n(B)$  for some  $(m, n)$  where either  $m > M$  or  $n > M$ , then  $f^{m_0} = g^{n_0}$ .

Analogous results have been proven in various cases in characteristic zero, in case  $E$  is replaced by  $\mathbb{A}^1$  [8], a linear space [6], or a semiabelian variety [6, 7]. Corollary 1.6, however, applies to all characteristics.

It would be interesting to study high-dimensional analogues of Corollary 1.6. For instance, we will show that if certain cases of Conjecture 1.4 hold, then Corollary 1.6 remains true if  $E$  is replaced by a *simple* abelian variety, i.e., an abelian variety having no nonzero proper abelian subvarieties. The referee's Proposition 1.5 thus yields an unconditional proof of the ineffective part of the simple abelian variety analogue of Corollary 1.6.

**Corollary 1.7.** *Let  $X$  be a simple abelian variety over an algebraically closed field  $k$ , and let  $f, g : X \rightarrow X$  be regular maps of degrees greater than 1. If there are points  $A, B \in X(k)$  such that the forward orbits  $O_f(A) := \{A, f(A), f^2(A), \dots\}$  and  $O_g(B) := \{B, g(B), g^2(B), \dots\}$  have infinite intersection, then  $f$  and  $g$  have a common iterate, namely,  $f^{m_0} = g^{n_0}$  for some positive integers  $m_0, n_0$ .*

The characteristic zero case of Corollary 1.6 and Corollary 1.7 is an instance of the higher-rank generalization posed in [8, Question 1.6] of the dynamical Mordell–Lang conjecture [2, Chapter 3]; see also [6]. For positive characteristic, see [2, Chapter 13]. We note that the conclusions of all previous results in characteristic  $p > 0$  involve the more complicated possibility of  $p$ -automatic sequences (e.g., [3, 5]), whereas the conclusion of Corollary 1.6 and Corollary 1.7 is more rigid. This extra possibility also occurs in the positive characteristic version of the original (not dynamical) Mordell–Lang conjecture [12], where it is called an “ $F$ -structure” and where examples are given to show that the possibility cannot be removed.

The rest of the paper is organized as follows. In Section 2, we state a known Diophantine result. Then Section 3, 4 and 5 contain proofs of Corollary 1.6 (together with Corollary 1.7), Theorem 1.2 and Theorem 1.3, respectively. The proofs are independent of one another, and can be read in any order. Theorem 1.1 follows immediately from Theorem 1.2 and Theorem 1.3. The appendix includes the referee's proof of Proposition 1.5 and a result about semigroups not containing noncommutative free semigroups.

## 2. LINEAR FORMS IN LOGARITHMS

The proofs of Theorem 1.2 and Theorem 1.3 rely on the following form of Baker's theorem on Diophantine approximation of logarithms.

**Theorem 2.1** (Baker, Wüstholz [1]). *Let  $\lambda_1, \dots, \lambda_r$  be complex numbers such that  $e^{\lambda_i}$  are algebraic for  $1 \leq i \leq r$ . Then there are effectively computable constants  $k, C > 0$  depending on  $r$  and  $\lambda_i$  such that*

$$0 < |a_1\lambda_1 + \cdots + a_r\lambda_r| \leq kH^{-C}$$

has no solutions in  $a_i \in \mathbb{Z}$ , where  $H = \max_{i=1}^r |a_i|$ .

The effective computability of Theorem 2.1 implies the effective part of our results, and our proofs will yield explicit bounds in our result if we use an explicit version of Theorem 2.1 (for example, see [4, §3.2]).

### 3. PROOF OF COROLLARY 1.6 AND COROLLARY 1.7

In this section, we prove Corollary 1.6 and a conditional generalization to simple abelian varieties, which implies Corollary 1.7.

*Proof of Corollary 1.6.* Since  $\deg(f) > 1$ , the regular map  $f$  has a fixed point. By replacing the origin of  $E$  by a fixed point of  $f$  if necessary, we may assume that  $f$  is an endomorphism of  $E$ .

Write  $g = \tau_Q \circ h$  where  $Q$  is a point on  $E$ ,  $\tau_Q$  is the map  $E \rightarrow E$  defined by translation by  $Q$ , and  $h$  is an endomorphism of  $E$ . Here  $\deg(h) = \deg(g) > 1$ , so that  $h - 1$  is nonconstant and thus induces a surjective map  $E \rightarrow E$ . Let  $R$  be a point on  $E$  such that  $(h - 1)(R) = Q$ . Then, for any positive integer  $n$ , we have

$$(3.1) \quad g^n = \tau_{Q+h(Q)+h^2(Q)+\cdots+h^{n-1}(Q)} \circ h^n = \tau_{(h^n-1)(R)} \circ h^n.$$

Thus, for any positive integer  $m$ , the condition  $f^m = g^n$  is equivalent to the conditions that  $f^m = h^n$  and  $(h^n - 1)(R) = O$ .

Pick the orbits of  $f$  and  $g$  that have infinite intersection, and let  $P$  be any point in the intersection; then the orbits  $O_f(P)$  and  $O_g(P)$  also have infinite intersection, so there are infinitely many pairs  $(m, n)$  of positive integers such that

$$f^m(P) = g^n(P) = (h^n - 1)(R) + h^n(P).$$

Fix such a pair  $(m_0, n_0)$ , and let  $(m, n)$  be any other pair of positive integers that satisfy the above. Then

$$(3.2) \quad (f^{m_0} - h^{n_0})(P) = (h^{n_0} - 1)(R)$$

$$(3.3) \quad (f^m - h^n)(P) = (h^n - 1)(R)$$

Left-multiplying (3.2) by the dual isogeny  $(\bar{h}^{n_0} - 1)$  of  $(h^{n_0} - 1)$ , we get

$$(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0})(P) = \deg(h^{n_0} - 1)(R)$$

Left-multiplying further by  $(h^n - 1)$ , we get

$$(h^n - 1)(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0})(P) = (h^n - 1) \deg(h^{n_0} - 1)(R)$$

Note that  $\deg(h^{n_0} - 1)$  is an integer, so it is in the center of  $\text{End}(E)$ . Using (3.3), we get

$$\left( (h^n - 1)(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0}) - (f^m - h^n) \deg(h^{n_0} - 1) \right) (P) = O.$$

Since  $O_f(P)$  is infinite,  $P$  must be a point of infinite order (otherwise,  $rP = 0$  for some integer  $r > 0$ , so  $O_f(P)$  lies in the finite group  $E[r]$  of  $r$ -torion elements).

Hence the kernel of  $(h^n - 1)(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0}) - (f^m - h^n) \deg(h^{n_0} - 1)$  contains all (infinitely many) multiples of  $P$ . Since the kernel of any nonzero endomorphism is a finite group, we must have

$$(3.4) \quad (h^n - 1)(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0}) - (f^m - h^n) \deg(h^{n_0} - 1) = 0$$

and recall that this holds for infinitely many pairs  $(m, n)$ .

Rewrite (3.4) as an equation in  $f^m$  and  $h^n$ :

$$(3.5) \quad h^n(u + d) - f^m d = u$$

where  $u = (\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0})$ ,  $d = \deg(h^{n_0} - 1)$ .

Now  $\text{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$  is either  $\mathbb{Q}$  or an imaginary quadratic field or a positive definite quaternion algebra over  $\mathbb{Q}$ , all of which can be embedded into some positive definite quaternion algebra  $H$  over  $\mathbb{Q}$ . View the equation (3.5) in  $H$ .

If  $u \neq 0$ , then the equation  $h^n(u + d)u^{-1} - f^m d u^{-1} = 1$  has infinitely many solutions  $m, n > 0$ , a contradiction to Theorem 1.1 with  $a = b = 1$ ,  $a' = (u + d)u^{-1}$ ,  $b' = -d u^{-1}$ ,  $\Gamma_1$  generated by  $h$ , and  $\Gamma_2$  generated by  $f$ . Hence  $u = 0$ , so that  $(\bar{h}^{n_0} - 1)(f^{m_0} - h^{n_0}) = 0$ .

But  $\deg \bar{h} = \deg h > 1$  implies  $\bar{h}^{n_0} - 1 \neq 0$ , so  $f^{m_0} = h^{n_0}$ .

Finally, equation (3.2) implies  $(h^{n_0} - 1)(R) = O$ , so  $g^{n_0} = h^{n_0} = f^{m_0}$  by (3.1). □

Let  $X$  be a simple abelian variety, and assume that Conjecture 1.4 holds for  $A = \text{End}(X) \otimes \mathbb{Q}$  and  $\Gamma_1, \Gamma_2$  being cyclic semigroups. We claim that Corollary 1.6 remains true if  $E$  is replaced by  $X$ . The proof is the same as above except for three places. First, we used the fact that  $h - 1$  is surjective because it is nonconstant. This is still true because the image of a morphism must be an abelian subvariety, but  $X$  is simple. Second, we used the elements  $\bar{h}^{n_0} - 1$ , but dual isogeny no longer exists in abelian varieties in general. However, we can fix an endomorphism  $\varphi$  such that  $\varphi \circ (h^{n_0} - 1) = \deg(h^{n_0} - 1)$ , and use  $\varphi$  in place of  $\bar{h}^{n_0} - 1$ . Third, we used the argument that if an endomorphism  $\psi$  of  $E$  vanishes at a point  $P$  of infinite order, then  $\psi = 0$ . This is also true for simple abelian variety  $X$ : the endomorphism  $\psi$  must vanish on the Zariski closure of the group generated by  $P$ , but it contains an abelian subvariety of  $X$  of positive dimension, which has to be the whole  $X$  because  $X$  is simple.

Given the referee's Proposition 1.5, the ineffective part of the conditional result above holds unconditionally, and it gives an unconditional proof of Corollary 1.7.

*Remark 3.1.* When  $k$  has characteristic zero, the ineffective part of Corollary 1.6 was proved via two different methods in [6, Theorem 1.4] and [13, Theorem 1.2.3]. Our proof of Corollary 1.6 extends the latter proof to arbitrary characteristic, and in fact the possibility of such an extension was the initial motivation for studying unit equations on quaternions in the present paper. We thank Michael Zieve for informing the author about [13, Theorem 1.2.3] and suggesting this possibility.

*Remark 3.2.* If  $f, g$  are endomorphisms of an elliptic curve  $E$  without translation, then Corollary 1.6 becomes trivial. For a proof, set  $P \in E(k)$  be a point in the intersection of orbits, and let  $n, m > 0$  be such that  $f^n(P) = g^m(P)$ . For any integer  $N$ , we have  $Nf^n(P) = Ng^m(P)$ , so that  $(f^n - g^m)(NP) = O$  because  $f, g$  are endomorphisms of  $E$ . But  $P$  is of infinite order (otherwise the forward orbit of  $P$  under  $f$  would be finite), so  $\ker(f^n - g^m)$  is an infinite group, and the only possibility is  $f^n - g^m = 0$ .

#### 4. PROOF OF THEOREM 1.2

Let  $\Delta$  be the set consisting of  $(f, g) \in \Gamma_1 \times \Gamma_2$  such that  $afa' + bgb' = 1$  and  $|1 - afa'| \neq |afa'|$ . Then the goal of Theorem 1.2 is precisely to show that  $\Delta$  is a finite set.

By triangle inequality, every  $(f, g) \in \Delta$  satisfies

$$(4.1) \quad 0 < \left| |afa'| - |bgb'| \right| \leq 1$$

We observe that since  $\Gamma_i$  ( $i = 1, 2$ ) is a semigroup generated by finitely many elements with norms greater than 1, there are only finitely many elements of  $\Gamma_i$  of bounded norm.

In the rest of the proof, we will prove the claim that  $\{|f| : (f, g) \in \Delta\}$  is bounded. Given the claim, the set  $\{f : (f, g) \in \Delta\}$  is finite by the observation above. Since  $f$  determines  $g$  by  $g = b^{-1}(1 - afa')b'^{-1}$ , there are only finitely many choices for  $g$  as well, and Theorem 1.2 is proved.

For contradiction, we assume that there is a solution  $(f, g) \in \Delta$  with arbitrarily large  $|f|$ . Using simple calculus (specifically, Lagrange's mean value theorem), (4.1) implies

$$(4.2) \quad 0 < \left| \log|afa'| - \log|bgb'| \right| \leq \frac{2}{|afa'|}$$

for sufficiently large  $|f|$ .

Let the semigroup  $\log|\Gamma_1|$  be generated by  $x_1, \dots, x_t > 0$  and  $\log|\Gamma_2|$  by  $y_1, \dots, y_u > 0$ . Write  $\log|f| = m_1x_1 + \dots + m_tx_t$ ,  $\log|g| = n_1y_1 + \dots + n_uy_u$  for some nonnegative integers  $m_i, n_j$ . Let  $c = \log|aa'/bb'|$ . Then  $c, x_i, y_j$  are logarithms of real algebraic numbers, and (4.2) can be rewritten as

$$(4.3) \quad 0 < |c + m_1x_1 + \dots + m_tx_t - n_1y_1 - \dots - n_uy_u| \leq \frac{2}{|a|e^{x_1m_1 + \dots + x_tm_t}}$$

By Theorem 2.1 (Baker's theorem), there are positive constants  $k, C$  such that

$$0 < |a_1c + m_1x_1 + \cdots + m_tx_t - n_1y_1 - \cdots - n_uy_u| \leq k \max\{|a_1|, |m_i|, |n_j|\}^{-C}$$

has no integer solution  $(a_1, m_1, \dots, m_t, n_1, \dots, n_u)$ . In particular, for  $a_1 = 1$  and  $m_i, n_j > 0$ , the inequality

$$(4.4)$$

$$0 < |c + m_1x_1 + \cdots + m_tx_t - n_1y_1 - \cdots - n_uy_u| \leq kH^{-C} \text{ has no solution,}$$

where  $H = \max\{1, m_1, \dots, m_t, n_1, \dots, n_u\}$ .

Our next goal is to bound the right-hand side of (4.3) by a function of  $H$ , in order to reach a contradiction with (4.4). Since  $x_i, y_j$  are positive, for  $|f|$  sufficiently large and satisfying (4.3), it is not hard to see that

$$(4.5) \quad C_1 \max\{m_i\} < \max\{n_j\} < C_2 \max\{m_i\}$$

for some  $C_1, C_2 > 0$  that does not depend on  $m_i, n_j$ . For a proof, we note that

$$\min\{x_i\} \max\{m_i\} \leq m_1x_1 + \cdots + m_tx_t \leq t \max\{x_i\} \max\{m_i\}$$

$$\min\{y_j\} \max\{n_j\} \leq n_1y_1 + \cdots + n_uy_u \leq u \max\{y_j\} \max\{n_j\}$$

and (4.3) gives

$$\frac{1}{2}(n_1y_1 + \cdots + n_uy_u) < m_1x_1 + \cdots + m_tx_t < 2(n_1y_1 + \cdots + n_uy_u)$$

for sufficiently large  $|f|$ . Hence  $\max\{m_i\}, \max\{n_j\}, \log |f|$  and  $\log |g|$  are all "comparable" to each other in the sense of (4.5)

It follows that

$$(4.6) \quad C_1 \max\{m_i\} < H \leq \max\{C_2, 1\} \max\{m_i\} =: C'_2 \max\{m_i\}$$

where we denote  $C'_2 = \max\{C_2, 1\}$ .

Now (4.3) implies

$$(4.7)$$

$$0 < |c + m_1x_1 + \cdots + m_tx_t - n_1y_1 - \cdots - n_uy_u| \leq \frac{2}{|a|e^{\min\{x_i\} \max m_i}} \leq \frac{2}{|a|e^{\min\{x_i\} H/C'_2}}$$

for sufficiently large  $|f|$  (or equivalently,  $H$ , by the "comparability" discussion above together with (4.6)).

Since the right-hand side decays exponentially in  $H$ , it will be less than  $kH^{-C}$  for large  $H$ , which contradicts the lack of solution of (4.4).  $\square$

## 5. PROOF OF THEOREM 1.3

First, we observe that the equation  $|1 - afa'| = |afa'|$  can be rewritten as  $|a^{-1}a'^{-1} - f| = |0 - f|$ . Note that  $|\cdot|$  is the norm induced from the inner product on  $\mathbb{H}$  with  $\{1, i, j, k\}$  being an orthonormal basis. We denote the inner product by  $\langle \cdot, \cdot \rangle$ .



Denoting  $d = a^{-1}a'^{-1}$ , the equation above gives

$$\begin{aligned}\langle f, f \rangle &= |f|^2 \\ &= |d - f|^2 \\ &= \langle d - f, d - f \rangle = |d|^2 - 2\langle d, f \rangle + |f|^2,\end{aligned}$$

which simplifies to  $2\langle d, f \rangle = |d|^2$ .

Hence the equation is equivalent to that  $f$  lies in a hyperplane not passing through the origin, given by

$$\{x \in \mathbb{H} : \langle a^{-1}a'^{-1}, x \rangle = \frac{1}{2}|a^{-1}a'^{-1}|^2\}.$$

Given the observation above, Theorem 1.3 follows from the following lemma:

**Lemma 5.1.** *Let  $\Gamma$  be a commutative semigroup of  $\mathbb{H}^\times$  generated by finitely many algebraic elements of norms greater than 1, and  $H$  be a hyperplane of  $\mathbb{H}$  defined by*

$$H = \{x \in \mathbb{H} : \Theta(x) = 1\}$$

where  $\Theta : \mathbb{H} \rightarrow \mathbb{R}$  is a nonzero  $\mathbb{R}$ -linear functional that maps  $\mathbb{H}_a$  into  $\overline{\mathbb{Q}} \cap \mathbb{R}$ . Then  $\Gamma \cap H$  is finite. In fact, we have an effectively computable upper bound that depends only on  $H$  and  $\Gamma$  for norms of elements of  $\Gamma \cap H$ .

*Proof of lemma.* Since  $\Gamma$  is commutative, it lies in a subalgebra in  $\mathbb{H}$  that is isomorphic to  $\mathbb{C}$ . Passing to its restriction on this subalgebra, we may assume instead that  $\Gamma$  is a semigroup generated by  $g_1, \dots, g_s \in \overline{\mathbb{Q}}^\times \subseteq \mathbb{C}$  such that  $|g_j| > 1$ , and  $\Theta : \mathbb{C} \rightarrow \mathbb{R}$  is an  $\mathbb{R}$ -linear functional (which could now be zero) that maps  $\overline{\mathbb{Q}}$  into  $\overline{\mathbb{Q}} \cap \mathbb{R}$ . We need to show that  $\Theta(f) = 1$  has only finitely many solutions  $f \in \Gamma$ .

There is no question to ask if  $\Theta = 0$ . In the case  $\Theta \neq 0$ , we may assume  $\Theta$  is given by  $\langle v, \cdot \rangle$  for some nonzero vector  $v \in \overline{\mathbb{Q}}$ , where  $\langle \cdot, \cdot \rangle$  is the standard Euclidean inner product on  $\mathbb{C}$  with  $\{1, i\}$  being an orthonormal basis. By rescaling, we may assume  $|v| = 1$ , but the equation  $\Theta(f) = 1$  will become

$$(5.1) \quad \langle v, f \rangle = M$$

for some real algebraic number  $M > 0$ .

Write  $g_j = r_j v_j$  with  $r_j > 1$  and  $v_j = e^{i\theta_j}$  on the unit circle, with  $0 \leq \theta_j < 2\pi$ . Also write  $v = e^{i\theta}$  with  $0 \leq \theta < 2\pi$ . For  $f = g_1^{n_1} \dots g_s^{n_s}$ , the equation (5.1) becomes

$$(5.2) \quad \langle v, e^{i(n_1\theta_1 + \dots + n_s\theta_s)} \rangle = Mr_1^{-n_1} \dots r_s^{-n_s}$$

The left-hand side involves the inner product of two unit vectors, so its value is  $\cos((n_1\theta_1 + \dots + n_s\theta_s) - \theta)$ . When  $n_i$  are sufficiently large, the right-hand side of 5.2 is small. But  $|\cos((n_1\theta_1 + \dots + n_s\theta_s) - \theta)|$  is approximately the closest distance from  $(n_1\theta_1 + \dots + n_s\theta_s) - \theta$  to  $(m + 1/2)\pi$  for integer

$m$ . If (5.2) is satisfied by infinitely many  $(n_j)$ 's, then for sufficiently large solutions  $(n_j)$ , we have

$$(5.3) \quad 0 < \left| \left( \frac{\pi}{2} + \theta \right) + m\pi - (n_1\theta_1 + \cdots + n_s\theta_s) \right| < 2Mr_1^{-n_1} \cdots r_s^{-n_s}$$

for some  $m \in \mathbb{Z}$ .

By assumption,  $v, v_j$  are algebraic numbers, so  $\lambda := i(\frac{1}{2}\pi + \theta)$ ,  $\mu = i\pi$  and  $\lambda_j = i\theta_j$  are logarithms of algebraic numbers. By Theorem 2.1, there are constants  $k, C > 0$  such that the inequality

$$(5.4) \quad 0 < \left| \left( \frac{\pi}{2} + \theta \right) + m\pi - (n_1\theta_1 + \cdots + n_s\theta_s) \right| < kB^{-C}$$

for  $m, n_j \in \mathbb{Z}, n_j \geq 0$ , where

$$B = \max\{1, |m|, n_j\}$$

But for solutions of (5.3) with  $n_j$  large,  $m\pi$  must be close to  $n_1\theta_1 + \cdots + n_s\theta_s - (\frac{1}{2}\pi + \theta)$ . Noting that

$$n_1\theta_1 + \cdots + n_s\theta_s \leq s \max\{\theta_j\} \max\{n_j\},$$

we have

$$(5.5) \quad |m| \leq C' \max\{n_j\}$$

for some constant  $C'$ , and thus

$$(5.6) \quad \max\{n_j\} \leq B = \max\{n_j, |m|\} \leq \max\{1, C'\} \max\{n_j\}$$

It follows from (5.4) that for some constant  $k' > 0$ ,

$$(5.7) \quad 0 < \left| \left( \frac{\pi}{2} + \theta \right) + m\pi - (n_1\theta_1 + \cdots + n_s\theta_s) \right| < k' \max\{n_j\}^{-C}$$

for  $m, n_j \in \mathbb{Z}, n_j \geq 0$ . But for  $(n_j)$  large,  $2Mr_1^{-n_1} \cdots r_s^{-n_s} < k' \max\{n_j\}^{-C}$ , yielding a contradiction with (5.3).  $\square$

## 6. FUTURE WORK

We were able to arrive at the main theorem using the archimedean norm only. If we can furthermore use some version of  $p$ -adic norm on the division algebra  $A$ , we can vastly improve the result by applying K. Yu's theorem about  $p$ -adic logarithms in [17]. One possible proposal for a  $p$ -adic norm is to use the reduced norm of a division algebra over  $\mathbb{Q}_p$ , which only works if  $A \otimes \mathbb{Q}_p$  is still a division algebra. Unfortunately, for each given  $A$ , this only holds for finitely many  $p$ .

Theorem 1.2 is potentially useful for more cases than in Theorem 1.1. For example, one can explore the analogue of Theorem 1.3 in the case where  $\Gamma$  has two or more noncommutative generators, and then apply Theorem 1.2. Even if  $\Gamma$  is replaced by its subset  $\{f_1^{n_1} f_2^{n_2} : n_1, n_2 \geq 0\}$ , where  $f_1, f_2$  are noncommutative generators with norms greater than 1, the analogue of Theorem 1.3 remains open.

The following example shows that we should only consider Conjecture 1.4 where  $A$  is a division algebra.

**Example 6.1.** Take  $A = M_2(\mathbb{Q})$ , the algebra of  $2 \times 2$  matrices over  $\mathbb{Q}$ . Then the multiplicative semigroup generated by  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  is

$$\Gamma := \left\{ \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} : n \in \mathbb{Z}, n \geq 0 \right\}.$$

The equation  $2f - g = 1_A$  has infinitely many solutions  $f, g \in \Gamma$ , namely all  $(f, g)$  with  $f \in \Gamma$  and  $g = f^2$ .

#### APPENDIX

This section contains the proofs of Proposition 1.5 and Proposition A.2, both sketched by the referee. We start with an observation that will be used in both proofs.

**Lemma A.1.** *Let  $A$  be a finite-dimensional division algebra over a field  $k$ , and let  $K$  be the algebraic closure of  $k$ . Then there is an embedding of  $k$ -algebras from  $A$  to the matrix algebra  $M_n(K)$  for some integer  $n > 0$ .*

*Proof.* Let  $L$  be the center of  $A$ . Then  $L$  is a finite extension of  $k$  and we can embed  $A$  into  $A \otimes_L K$ , which is isomorphic to the matrix algebra  $M_n(K)$  for some integer  $n$  by a standard fact about central simple algebras.  $\square$

*Proof of Proposition 1.5.* Let  $K$  be the algebraic closure of  $k$  and fix an embedding  $A \hookrightarrow M_n(K)$  as in Lemma A.1. Note that nonzero elements of  $A$  are sent to invertible matrices in  $M_n(K)$ . From now on, we shall consider the unit equation in  $M_n(K)$ .

To set up a proof by contradiction, we assume that

$$(A.1) \quad a_1 f_1 b_1 + \cdots + a_m f_m b_m = 1, (f_1, \dots, f_m) \in \Gamma_1 \times \cdots \times \Gamma_m$$

is a shortest equation (i.e., with minimal  $m$ ) in the setting of Proposition 1.5 that has infinitely many degenerate solutions. We claim:

There cannot exist an infinite family of degenerate solutions

$$(A.2) \quad \{(f_1^\alpha, \dots, f_m^\alpha)\}$$

indexed by  $\alpha$  in an infinite set, such that  $f_1^\alpha$  are the same for all  $\alpha$ .

Otherwise, call  $f_1^\alpha = f_1$ , and let  $u = a_1 f_1 b_1$ , which is not 1 because the solution is nondegenerate. Then  $1 - u$  is a unit in  $A$  because  $A$  is a division algebra, and set  $b'_i = b_i(1 - u)^{-1}$ . The following equation

$$a_2 f_2 b'_2 + \cdots + a_m f_m b'_m = 1$$

has infinitely many nondegenerate solutions  $(f_2, \dots, f_m) = (f_2^\alpha, \dots, f_m^\alpha)$ , contradicting the minimality of  $m$ .

Now note that every element  $\gamma$  of  $\Gamma_i$  is diagonalizable in  $M_n(K)$ . Indeed, since  $\gamma \in R$  and  $R$  is finite-dimensional over  $k$ , we see that  $\gamma$  satisfies some

minimal polynomial  $p(\gamma) = 0$  where  $p(x) \in k[x]$  is monic. To show that  $\gamma$  is diagonalizable, it suffices to show that  $p(x)$  has no repeated root in  $K$ . Assume the contrary, then there is a proper divisor  $p_0(x) \in k[x]$  of  $p(x)$  such that  $p(x)$  divides  $p_0(x)^2$ . Thus  $p_0(\gamma)^2 = 0$ , so that  $p_0(\gamma) = 0$  because  $R$  is a division algebra. This is a contradiction to the minimality of  $p(x)$ .

Since  $\Gamma_i$  is abelian and finitely generated, and every element of  $\Gamma_i$  is diagonalizable, there is a simultaneous diagonalization of  $\Gamma_i$  by some  $s_i \in GL_n(K)$ , i.e.,  $s_i\Gamma_i s_i^{-1}$  only consists of diagonal matrices in  $M_n(K)$ . So we may replace  $a_i$  by  $a_i s_i^{-1}$ ,  $b_i$  by  $s_i b_i$ , and  $\Gamma_i$  by  $s_i\Gamma_i s_i^{-1}$  and assume that each  $\Gamma_i$  only consists of diagonal matrices in  $M_n(K)$  (though  $\Gamma_i, a_i, b_i$  are no longer inside  $R$ , but it will not matter).

Now consider a solution  $(f_i)$  of  $1 = a_1 f_1 b_1 + \cdots + a_m f_m b_m$ , where  $f_i = \text{diag}(x_{i1}, \dots, x_{in})$ . Looking at the  $(1, 1)$ -entry, we obtain an equation of the form

$$(A.3) \quad 1 = \sum_{1 \leq i, j \leq n} p_{ij} x_{ij}$$

for some fixed  $p_{ij} \in K$ . (To see it, one merely needs to notice that every entry of  $a_1 f_1 b_1 + \cdots + a_m f_m b_m$  is a linear combination of entries of  $f_i$ .)

Let  $S$  be the set consisting of all  $(i, j)$  such that  $p_{ij} \neq 0$ . We may assume that  $S$  is nonempty, otherwise (A.3) has no solution and there is nothing to prove. We claim that

$$(A.4) \quad \text{There are finite sets } X_{ij} \text{ for } (i, j) \in S, \text{ such that whenever } (x_{ij}) \text{ is a solution of (A.3), there exists } (i_0, j_0) \in S \text{ such that } x_{i_0 j_0} \in X_{i_0 j_0}.$$

To prove the claim, notice that there is a finitely generated subgroup of  $K^\times$  that contains all  $x_{ij}$  because  $\Gamma_i$  is finitely generated. By the  $S$ -unit theorem in several variables [4, Theorem 6.1.3], for every nonempty subset  $T$  of  $S$ , the equation

$$(A.5) \quad 1 = \sum_{(i,j) \in T} p_{ij} x_{ij}$$

has only finitely many nondegenerate solutions. Let  $X_{ij}$  be the set consisting of all  $x_{ij}$  that appears in a nondegenerate solution of (A.5) for some  $T$ . Now for every solution  $(x_{ij})$  of (A.3), there is a minimal nonempty subset  $T_0$  of  $S$  such that  $1 = \sum_{(i,j) \in T_0} p_{ij} x_{ij}$ . Pick  $(i_0, j_0) \in T_0$ , then by construction,  $x_{i_0 j_0} \in X_{i_0 j_0}$ , as required.

Now applying the pigeonhole principle to the infinitely many solutions of (A.1), we see that there exists  $(i_0, j_0) \in S$  and  $x \in X_{i_0 j_0}$  such that there are infinitely many solutions  $(f_i) = (\text{diag}(x_{ij}))$  with  $x_{i_0 j_0} = x$ . Without loss of generality, assume  $(i_0, j_0) = (1, 1)$ . We claim that all those solutions  $(f_i)$  have the same  $f_1$ . This contradicts (A.2) and finishes the proof of Proposition 1.5.

It remains to prove the claim. Let  $(f_i) = \text{diag}(x_{i1}, \dots, x_{in})$  and  $(f'_i) = \text{diag}(x'_{i1}, \dots, x'_{in})$  be two solutions of (A.1) such that  $x_{11} = x'_{11} = x$ . Then

$g := f_1(f'_1)^{-1}$  is a diagonal matrix with  $(1, 1)$ -entry being 1. In particular,  $g - 1$  is not invertible in  $M_n(K)$ . But by construction,  $h := s_1^{-1}(g - 1)s_1$  is in  $A$ , so  $h$  is either zero or a unit. Since  $g - 1$  is not invertible, neither is  $h$ . Therefore  $h = 0$ , so that  $g = 1$ , which gives  $f_1 = f'_1$ .  $\square$

We next prove a statement that implies a generalization of Proposition 1.5 where all  $\Gamma_i$  are semigroups not containing free semigroups of rank two. Recall that a group  $G$  is *virtually  $P$*  (where  $P$  is a property) if  $G$  has a finite-index subgroup that is  $P$ .

**Proposition A.2.** *Let  $A$  be a division algebra over a field  $k$ , and  $\Gamma$  be a finitely generated semigroup of  $A^\times$ . If  $\Gamma$  does not contain a free semigroup of rank two, then the group  $G$  generated by  $\Gamma$  in  $A^\times$  is virtually abelian.*

*Proof.* First, we will show that  $G$  is virtually nilpotent using Theorem 1 of [14]. Embed  $A$  into  $M_n(K)$  as in Lemma A.1, where  $K$  is the algebraic closure of  $k$ . Note that  $\Gamma$  is in  $GL_n(K)$ .

Since  $\Gamma$  is finitely generated, there is a finitely generated subfield  $L \subseteq K$  such that  $\Gamma \subseteq GL_n(L)$  (for example, let  $L$  be the field generated by matrix entries of generators of  $\Gamma$  over the prime field  $\mathbb{Q}$  or  $\mathbb{F}_p$  of  $k$ ). Now, Theorem 1 of [14] implies that  $G$  is virtually nilpotent.

Let  $N$  be a finite-index nilpotent subgroup of  $G$ . By the following lemma,  $N$  is virtually abelian, and the proof is complete.  $\square$

*Warning.* Here  $L$  may not contain  $k$ , but it does not matter for the purpose of this proof.

**Lemma A.3.** *If  $A$  is a division algebra over a field  $k$ , and  $N \subseteq A^\times$  is a solvable subgroup, then  $N$  is virtually abelian.*

*Proof.* Again, we embed  $A$  into  $M_n(K)$  as in Lemma A.1, where  $K$  is the algebraic closure of  $k$ . Then  $N$  is a subgroup of  $GL_n(K)$ . The Zariski closure  $\overline{N}$  of  $N$  in  $GL_n(K)$  is a  $K$ -algebraic group that is still solvable, and so is its identity component  $\overline{N}_0$ . Let  $N_0 := \overline{N}_0 \cap N$ . Since  $\overline{N}_0$  has finite index in  $\overline{N}$ , so does  $N_0$  in  $N$ .

We claim that  $N_0$  is abelian. By the Lie–Kolchin triangularization theorem [15, Theorem 6.3.1], there is  $s \in GL_n(K)$  such that  $s\overline{N}_0s^{-1}$  consists of upper triangular matrices, so  $sN_0s^{-1}$  does as well. We observe that if  $a$  and  $b$  are invertible upper triangular matrices, then the commutator  $[a, b] := aba^{-1}b^{-1}$  is in  $U$ , the group of upper triangular matrices with diagonal entries all 1. It follows that  $s[N_0, N_0]s^{-1} \subseteq U$ .

It remains to show that  $[N_0, N_0]$  is the trivial group. Take  $x \in [N_0, N_0]$ , and note that  $x \in A$ , so  $x - 1 \in A$  is either zero or invertible. But  $s(x - 1)s^{-1} = xs^{-1} - 1$  is an upper triangular matrix in  $Mat_n(K)$  with diagonal entries all 0, so  $x - 1$  cannot be invertible. It follows that  $x = 1$  and  $N_0$  is abelian.  $\square$

## REFERENCES

- [1] A. Baker and G. Wüstholz. Logarithmic forms and group varieties. *J. Reine Angew. Math.*, 442:19–62, 1993.
- [2] J. P. Bell, D. Ghioca, and T. J. Tucker. *The Dynamical Mordell–Lang Conjecture*. American Mathematical Society, Providence, RI, 2016.
- [3] H. Derksen. A Skolem–Mahler–Lech theorem in positive characteristic and finite automata. *Invent. Math.*, 168(1):175–224, 2007.
- [4] J.-H. Evertse and K. Györy. *Unit Equations in Diophantine Number Theory*. Cambridge University Press, Cambridge, 2015.
- [5] D. Ghioca. The dynamical Mordell–Lang conjecture in positive characteristic. *Trans. Amer. Math. Soc.*, 371(2):1151–1167, 2019.
- [6] D. Ghioca and K. D. Nguyen. The orbit intersection problem for linear spaces and semiabelian varieties. *Math. Res. Lett.*, 24(5):1263–1283, 2017.
- [7] D. Ghioca, T. Tucker, and M. E. Zieve. The Mordell–Lang question for endomorphisms of semiabelian varieties. *J. Théor. Nombres Bordeaux*, 23(3):645–666, 2011.
- [8] D. Ghioca, T. J. Tucker, and M. E. Zieve. Linear relations between polynomial orbits. *Duke Math. J.*, 161(7):1379–1410, 2012.
- [9] M. Hindry and J. H. Silverman. *Diophantine Geometry: an Introduction*. Springer-Verlag, New York, 2013.
- [10] P. Koymans and C. Pagano. On the equation  $X_1 + X_2 = 1$  in finitely generated multiplicative groups in positive characteristic. *Q. J. Math.*, 68(3):923–934, 2017.
- [11] S. Lang. Integral points on curves. *Publications Mathématiques de l’IHÉS*, 6:27–43, 1960.
- [12] R. Moosa and T. Scanlon.  $F$ -structures and integral points on semiabelian varieties over finite fields. *Amer. J. Math.*, 126(3):473–522, 2004.
- [13] A. Odesky. *The orbit intersection problem and polynomial functions*. PhD Thesis, the University of Michigan–Ann Arbor, 2020.
- [14] J. Okniński and A. Salwa. Generalised Tits alternative for linear semigroups. *J. Pure Appl. Algebra*, 103(2):211–220, 1995.
- [15] T. A. Springer. *Linear algebraic groups*. Second edition. Birkhäuser Boston, Inc., Boston, MA, 1998.
- [16] J. F. Voloch. The equation  $ax + by = 1$  in characteristic  $p$ . *J. Number Theory*, 73(2):195–200, 1998.
- [17] K. Yu.  $p$ -adic logarithmic forms and group varieties III. *Forum Math.*, 19(2):187–280, 2007.