Hypergeometric dispersion and the orbit problem

Sergei A. Abramov*
Computer Center of the
Russian Academy of Science
Vavilova 40, Moscow 117967, Russia
abramov@ccas.ru

Manuel Bronstein
INRIA – Projet Café
2004, Route des Lucioles, B.P. 93
F-06902 Sophia Antipolis Cedex, France
Manuel.Bronstein@sophia.inria.fr

ABSTRACT

We describe an algorithm for finding the positive integer solutions n of *orbit problems* of the form $\alpha^n = \beta$, where α and β are given elements of a field K. Our algorithm corrects the bounds given in [7], and shows that the problem is not polynomial in the Euclidean norms of the polynomials involved. Combined with a simplified version of the algorithm of [8] for the "specification of equivalence", this yields a complete algorithm for computing the dispersion of polynomials in nested hypergeometric extensions of rational function fields. This is a necessary step in computing symbolic sums, or solving difference equations, with coefficients in such fields. We also solve the related equations $p(\alpha^n) = 0$ and $p(n, \alpha^n) = 0$ where p is a given polynomial and α is given.

1. INTRODUCTION

Given a polynomial ring $K[X]$ over a field K and an automorphism σ of $K[X]$, the *dispersion (w.r.t. σ)* is the function

$$\mathrm{Dis}_\sigma : K[X] \times K[X] \to \mathbb{Z} \cup \{+\infty\}$$

given by

$$\mathrm{Dis}_\sigma(p, q) = \max\{n \geq 0 \text{ such that } \deg(\gcd(p, \sigma^n q)) > 0\},$$

where $\max(\emptyset) = -1$ by convention. We also write $\mathrm{Dis}_\sigma(p)$ for $\mathrm{Dis}_\sigma(p, p)$. Introduced with respect to the shift $\sigma X = X + 1$ in [1] in order to compute rational sums, this function plays a key role in symbolic summation algorithms [1, 8, 9] as well as for solving linear ordinary difference and q–difference equations [3, 5]. We call it the *hypergeometric dispersion* when X is hypergeometric over K, *i.e.* $\sigma(X)/X \in K$. When σ is the identity on K and either $\sigma X = X + 1$ or $\sigma(X)/X \in K$, then it can be computed by looking at the largest integer root of $\mathrm{res}_X(p, \sigma^m q) \in K[m]$, where res_X denotes the resultant operation in $K[X]$. We address in this paper the problem of computing it in the more general case when X is hypergeometric over K and K is a tower of nested hypergeometric extension of C (resp. $C(n)$), where

$$C = \{a \in K \text{ such that } \sigma a = a\}$$

is the invariant field of K (resp. and $\sigma n = n + 1$).

The, seemingly unrelated, *σ–orbit problem* is, given nonzero α, β in a field F and an automorphism σ of F, to compute all the integers $n > 0$ such that $\prod_{i=0}^{n-1} \sigma^i \alpha = \beta$. In the base case, when σ is the identity on F and the equation to solve is $\alpha^n = \beta$, this problem is known as the *orbit problem*. It has many important applications in theoretical computer science since many reachability problems can be reduced to it: the accessibility problem for linear sequential machines [6], whether a power of a given matrix equals another given matrix, whether a given vector is reached from another by iterating a given linear map etc. A solution in the base case and when F is an algebraic number field was presented in [7].

The link between the dispersion and the orbit problem was provided by Karr, who reduced in [8] the computation of the dispersion over $\Pi\Sigma$–fields, which include hypergeometric extensions, to solving the orbit problem in the invariant field of K. The orbit problem was not yet solved at that time, so he made the additional hypothesis of an existence of an algorithm for solving the orbit problem in that field. One might conclude that combining the algorithms of [8] and [7] solves the dispersion problem, but it turns out that the bound presented in [7] is incorrect when $|\alpha| = |\beta| = 1$, and that in that case the solution of the orbit problem is not bounded by any polynomial of the Euclidean lengths of the inputs as claimed in [7].

We describe in this paper a complete algorithm for computing dispersions in nested hypergeometric extensions, by combining Karr's reduction with correct bounds for the orbit problem. We also discuss the related *power root* problem as well as a bivariate variant that occurs when computing hypergeometric dispersions without factorization, using resultants instead.

Notations and conventions: Given a ring R, any map $\sigma : R \to R$, $x \in R$ and a positive integer n, we let

$$x^{n,\sigma} = \prod_{i=0}^{n-1} \sigma^i x.$$

All rings and fields are of characteristic 0.

*Supported by a grant from the French–Russian Lyapunov Institute, Project No 98-03.

2. KARR'S REDUCTIONS

We describe in this section a restricted version of Karr's algorithm that is sufficient to compute dispersions in nested σ–hypergeometric extensions. Following [9], we say that F is Π*–regular w.r.t.* σ if for any $a \in F$ and $n > 0$,

$$a^{n,\sigma} = 1 \implies a^n = 1\,.$$

For an example of a non Π–regular field, consider $\mathbb{Q}(X)$ with σ the automorphism that maps X to $1-X$. Then,

$$\left(\frac{X}{1-X}\right)^{2,\sigma} = \frac{X}{1-X}\frac{1-X}{X} = 1$$

although $X/(1-X)$ is not a root of unity. Note that when $\alpha \in F^*$ is not a root of unity and F is Π–regular, then the σ–orbit problem has at most one solution. Indeed, if $\alpha^{n,\sigma} = \alpha^{m,\sigma}$ for $m > n > 0$, then

$$1 = \prod_{i=n}^{m-1} \sigma^i \alpha = \sigma^n\left(\alpha^{m-n,\sigma}\right),$$

which implies that $\alpha^{m-n,\sigma} = 1$, hence that $\alpha^{m-n} = 1$. We then say that F is *computably* Π*–regular w.r.t.* σ if there is an algorithm for solving the σ–orbit problem when α is not a root of unity. We also say that $\alpha \in F$ is a σ*–radical* over a subfield K of F if $\sigma x = \alpha^n x$ for some $x \in K^*$ and an integer $n > 0$. Note that roots of unity are σ–radicals over any subfield of F, since $\sigma 1 = 1 = \alpha^m 1$ when $\alpha^m = 1$. We can now describe Karr's first algorithm, which reduces computing the dispersion of irreducibles in $K[X]$ to solving σ–orbit problems over K.

THEOREM 1. *[8] Let $K[X]$ be a polynomial ring over a field K and σ be an automorphism of $K[X]$ mapping K onto K and X to aX for some $a \in K^*$. If a is not a σ–radical over K and K is computably Π–regular, then there is an algorithm for computing $Dis_\sigma(p,q)$ for any irreducible $p, q \in K[X]$.*

PROOF. The proof is contained in the proof of Theorem 4 of [8], but is reproduced here because it contains the algorithm. Let $p, q \in K[X]$ be irreducible and suppose that $\deg(\gcd(p, \sigma^m q)) > 0$ for some integer $m > 0$. Since σ preserves the degree and maps irreducibles to irreducibles, it follows that $\sigma^m q = up$ for some $u \in K^*$ and that $\deg(p) = \deg(q)$. Since $\sigma X = aX$, $\sigma^n X = a^{n,\sigma} X$ for $n \ge 0$. Write $p = \sum_{i=0}^d p_i X^i$ and $q = \sum_{i=0}^d q_i X^i$ where $p_d \neq 0 \neq q_d$. Then,

$$\sigma^m q = \sum_{i=0}^d \sigma^m q_i \sigma^m X^i = \sum_{i=0}^d \sigma^m(q_i)(a^{m,\sigma})^i X^i = up\,,$$

which implies that $\sigma^m(q_i)(a^{m,\sigma})^i = up_i$ for $0 \le i \le d$, hence that $p_i \neq 0 \iff q_i \neq 0$ and

$$\begin{aligned}\left(a^{d-i}\frac{\sigma(q_d/q_i)}{q_d/q_i}\right)^{m,\sigma} &= (a^{m,\sigma})^{d-i}\frac{\sigma^m(q_d/q_i)}{q_d/q_i} \\ &= \frac{p_d/p_i}{\sigma^m(q_d/q_i)}\frac{\sigma^m(q_d/q_i)}{q_d/q_i} = \frac{p_d/p_i}{q_d/q_i}\end{aligned}$$

for $0 \le i < d$ such that $p_i \neq 0 \neq q_i$. If $p_i = q_i = 0$ for $0 \le i < d$, then $d = 1$, $p = p_1 X$ and $q = q_1 X$, implying that $\mathrm{Dis}_\sigma(p,q) = +\infty$. Otherwise, let $i < d$ be such that $p_i \neq 0 \neq q_i$. Then, $\deg(\gcd(p, \sigma^m q)) > 0$ if and only if $\alpha^{m,\sigma} = \beta$ where $\beta = p_d q_i / p_i q_d \in K^*$ and

$$\alpha = a^{d-i}\sigma(q_d/q_i)/(q_d/q_i) \in K^*\,.$$

Since a is not a σ–radical over K, α is not a root of unity, so $\mathrm{Dis}_\sigma(p,q)$ can be computed. Note that it is finite. □

Although the above theorem requires p and q to be irreducible, if we have an algorithm for factoring the elements of $K[X]$ into irreducibles, Theorem 7 of [5] shows how computing dispersions of arbitrary polynomials can be reduced to computing dispersions of irreducibles. We are thus reduced to solving σ–orbit problems in K. If σ is the identity on K, then this is the classical orbit problem in K. Otherwise, if K is the fraction field of a polynomial ring $F[Y]$ where F is a subfield of K, then Karr's second reduction, described below, reduces the σ–orbit problem in K to computing dispersions in $F[Y]$.

We first need to extend the notion of dispersion to fractions: if σ is an automorphism of $F[Y]$, then for $p, q \in F[Y]$ such that $q \neq 0$ and $\gcd(p,q) = 1$, we define

$$\mathrm{Dis}_\sigma\left(\frac{p}{q}\right) = \max\{\mathrm{Dis}_\sigma(p), \mathrm{Dis}_\sigma(p,q), \mathrm{Dis}_\sigma(q,p), \mathrm{Dis}_\sigma(q)\}\,.$$

Recall that the *order at* ∞ is the function $\nu_\infty(q) = -\deg(q)$ for $q \in F[Y] \setminus \{0\}$, and given an irreducible $p \in F[Y]$, the *order at* p is the function

$$\nu_p(q) = \max\{n \in \mathbb{Z} \text{ such that } p^n | q\}$$

for $q \in F[Y] \setminus \{0\}$. They satisfy $\nu_\infty(ab) = \nu_\infty(a) + \nu_\infty(b)$ and $\nu_p(ab) = \nu_p(a) + \nu_p(b)$ for $a, b \in F[Y] \setminus \{0\}$. Both functions are extended to fractions via $\nu_\infty(a/b) = \nu_\infty(a) - \nu_\infty(b)$ and $\nu_p(a/b) = \nu_p(a) - \nu_p(b)$. Karr's algorithm relies on the following properties of those functions.

LEMMA 1. *Let $F[Y]$ be a polynomial ring over a field F and σ be an automorphism of $F[Y]$ mapping F onto F and Y to $aY + b$ for some $a \in F^*$ and $b \in F$. Then, for any $f \in F(Y)^*$ and any integer $n > 0$,*

(i) $\nu_\infty(f^{n,\sigma}) = n\nu_\infty(f)$.

(ii) $\nu_p(f^{n,\sigma}) = n\nu_p(f)$ *for any irreducible* $p \in F[Y]$ *such that* $p \mid \sigma p$.

(iii) $f \notin F \implies Dis_\sigma(f^{n,\sigma}) = Dis_\sigma(f) + n - 1$.

PROOF. (i) (ii) Write $f = q/h$ where $q, h \in F[Y] \setminus \{0\}$ and $\gcd(q,h) = 1$. Since σ is a morphism, $f^{n,\sigma} = (q^{n,\sigma})/(h^{n,\sigma})$. Since σ preserves the degree, $\deg(r^{n,\sigma}) = n\deg(r)$ for any $r \in F[Y] \setminus \{0\}$, so $\nu_\infty(f^{n,\sigma}) = n\deg(h) - n\deg(q) = n\nu_\infty(f)$. Since $p \mid \sigma p$ and σ maps irreducibles to irreducibles, $\sigma p = up$ for some $u \in F^*$. Let $r \in F[Y] \setminus \{0\}$ and write $r = p^\nu t$ where $\nu = \nu_p(r)$ and $t \in F[Y]$ is such that p does not divide t. Then, $up = \sigma p$ does not divide σt, so p does not divide σt. Therefore, $\sigma r = \sigma(p)^\nu \sigma t = u^\nu p^\nu \sigma t$ so $\nu_p(\sigma r) = \nu = \nu_p(r)$, which implies that $\nu_p(f^{n,\sigma}) = n\nu_p(q) - n\nu_p(h) = n\nu_p(f)$.

(iii) For any irreducible $p \in F[Y]$, we say that $g \in F(Y)^*$ is p*–orbital* if g can be written as $g = u\prod_{i=\alpha}^{\beta} \sigma^i(p)^{e_i}$ where

$u \in F^*$ and the e_i's are integers such that and $e_\alpha e_\beta \neq 0$. An *orbital decomposition* of $g \in F(Y)^*$ is a factorization $g = g_1 \dots g_m$ such that each g_i is p_i–orbital for some irreducible p_i and $\mathrm{Dis}_\sigma(p_i, p_j) = -1$ for $i \neq j$. Suppose first that $f \in F(Y) \setminus F$ is p–orbital and write $f = u \prod_{i=\alpha}^{\beta} \sigma^i(p)^{e_i}$. Then,

$$f^{n,\sigma} = (u^{n,\sigma}) \prod_{i=\alpha}^{\beta+n-1} \sigma^i(p)^{f_i}$$

where $f_\alpha = e_\alpha \neq 0$ and $f_\beta = e_\beta \neq 0$. If $p \mid \sigma^m p$ for some integer $m > 0$, then $\mathrm{Dis}_\sigma(f) = \mathrm{Dis}_\sigma(f^{n,\sigma}) = +\infty$ by Theorem 6 of [5]. Otherwise, Lemma 17 of [5] states that $\mathrm{Dis}_\sigma(f) = \beta - \alpha$ and that $\mathrm{Dis}_\sigma(f^{n,\sigma}) = \beta + n - 1 - \alpha = \mathrm{Dis}_\sigma(f) + n - 1$.

Let now $f \in F(Y) \setminus F$ be arbitrary. By Lemma 17 of [5], f has an orbital decomposition $f = f_1 \dots f_m$. It follows that $f^{n,\sigma} = ({f_1}^{n,\sigma}) \dots ({f_m}^{n,\sigma})$ is an orbital decomposition of $f^{n,\sigma}$, and Lemma 15 of [5] implies that

$$\begin{aligned} \mathrm{Dis}_\sigma(f^{n,\sigma}) &= \max_{1 \le i \le n} (\mathrm{Dis}_\sigma({f_i}^{n,\sigma})) \\ &= \max_{1 \le i \le n} (\mathrm{Dis}_\sigma(f_i) + n - 1) \\ &= \max_{1 \le i \le n} (\mathrm{Dis}_\sigma(f_i)) + n - 1 = \mathrm{Dis}_\sigma(f) + n - 1\,. \end{aligned}$$

□

We need some additional terminology: $p \in F[Y]$ is called *semi–invariant (w.r.t. σ)* if $\sigma(p)/p \in F^*$, and it is called *semi–periodic (w.r.t. σ)* if $\sigma^m(p)/p \in F^*$ for some integer $m > 0$. The smallest such m is then called the *period* of p (w.r.t. σ). Karr's reduction of the σ–orbit problem to dispersions now follows.

THEOREM 2. *[8] Let $F[Y]$ be a polynomial ring over a field F, σ be an automorphism of $F[Y]$ mapping F onto F and Y to $aY + b$ for some $a \in F^*$ and $b \in F$. If every semi–periodic $p \in F[Y]$ has period 1, if F is computably Π–regular and if there is an algorithm for computing $Dis_\sigma(p, q)$ for any $p, q \in F[Y]$, then $F(Y)$ is computably Π–regular.*

PROOF. This is essentially Theorem 5 of [8], with the hypothesis that $F[Y]$ is a $\Pi\Sigma$–extension of F replaced by the weaker hypothesis that every semi–periodic polynomial is a semi–invariant. Let $f, g \in F(Y)^*$ and suppose that f is not a root of unity and that $f^{n,\sigma} = g$ for some $n > 0$. If $\nu_\infty(f) \neq 0$, then $n = \nu_\infty(g)/\nu_\infty(f)$ by part (i) of Lemma 1. If $\nu_p(f) \neq 0$ for any irreducible semi–periodic $p \in F[Y]$, then $n = \nu_p(g)/\nu_p(f)$ by part (ii) of Lemma 1. Otherwise, neither the numerator nor the denominator of f has a semi–periodic factor, and Theorem 6 of [5] implies that $\mathrm{Dis}_\sigma(f)$ is finite. If $f \notin F$, then $n = \mathrm{Dis}_\sigma(g) - \mathrm{Dis}_\sigma(f) + 1$ by part (iii) of Lemma 1. Otherwise, $f \in F$, which implies that $g \in F$ and the value of n can be computed since F is computably Π–regular. □

We can now apply the above reductions one after the other to compute dispersions in towers of σ–hypergeometric extensions: let $R = C(X_1, \dots, X_m)[X_{m+1}]$ (respectively $C(X_0)$ $(X_1, \dots, X_m)[X_{m+1}]$) where σ is the identity on C, $\sigma X_0 = X_0 + 1$ and $\sigma(X_{i+1})/X_{i+1} = a_i \in K_i^*$ for $0 \le i \le m$, where $K_i = C(X_1, \dots, X_i)$ (resp. $C(X_0)(X_1, \dots, X_i)$). Suppose that a_i is not a σ–radical over K_i for $0 \le i \le m$, which implies that the hypotheses of Theorems 1 and 2 are satisfied throughout the tower (this follows from Theorem 2 of [8] or Theorem 4 of [5]). Theorem 1 reduces computing dispersions in R to solving σ–orbit problems in K_m. This is reduced by Theorem 2 to computing dispersions in $K_{m-1}[X_m]$, which is again reduced by Theorem 1 to solving σ–orbit problems in K_{m-1}. Continuing this process, we eventually get to solving σ–orbit problems in either C or $C(X_0)$. In C, this is the classical orbit problem, which we address in the next section. In $C(X_0)$, this is reduced by Theorem 2 to computing dispersions in $C[X_0]$, a problem which is classically solved by computing the integer roots of $\mathrm{res}_{X_0}(p(X_0), q(X_0 + m)) \in C[m]$, although one can also use factorization into irreducibles and Theorem 7 of [5].

3. THE ORBIT PROBLEM

We now turn to the orbit problem, *i.e.* given α, β in a field F, find the positive integer solutions of $\alpha^n = \beta$. To avoid the various open problems associated with transcendental numbers, we make the following computability assumptions about F: that we can test whether an element of F is algebraic or transcendental over $\mathbb{Q}$, and that we can test whether any two elements of F are algebraically independent over $\mathbb{Q}$. Suppose first that α is transcendental over $\mathbb{Q}$. If α and β are algebraically independent over $\mathbb{Q}$, then the corresponding orbit problem has no solution. Otherwise, β is algebraic over $\mathbb{Q}(\alpha)$ and looking at the degree at which α appears in β gives at most one candidate solution for the orbit problem, which is then solved. So suppose from now on that α is algebraic over $\mathbb{Q}$ and let $f_\alpha \in \mathbb{Q}[X]$ be its monic minimal polynomial over $\mathbb{Q}$ and $n_\alpha = \deg(f_\alpha)$. Since the orbit problem has no solution if β is transcendental over $\mathbb{Q}$ or if $\beta \notin \mathbb{Q}(\alpha)$ (something that we can test when β is algebraic over $\mathbb{Q}$), we can assume that $\beta \in \mathbb{Q}(\alpha)$, hence that $\beta = q(\alpha)$ for some polynomial $q \in \mathbb{Q}[X]$ such that $\deg(q) < n_\alpha$. If $\alpha^m = 1$ for some integer $m > 0$, then we can test whether $\alpha^i = \beta$ for $0 \le i < m$. If this is not the case, then the orbit problem has no solution, otherwise its solutions consist of all the integers of the form $i_0 + km_0$ where $k \ge 0$, i_0 is the smallest $i \ge 0$ such that $\alpha^i = \beta$ and m_0 is the smallest $m > 0$ such that $\alpha^m = 1$. So we assume in the rest of this section that α is not a root of unity, which implies that the orbit problem has at most one solution. We first give a counterexample to Theorem 3 of [7], in which the solution m is not bounded by any polynomial in n_α, $\log|q|$ and $\log|f_\alpha|$, where $|p|$ denotes the Euclidean norm of the vector of coefficients of $p \in \mathbb{Q}[X]$.

EXAMPLE 1. *Take $\alpha = (3 + 4\sqrt{-1})/5$, then $f_\alpha(X) = X^2 - 6X/5 + 1$, $n_\alpha = 2$ and $|f_\alpha| = \sqrt{1 + 36/25 + 1} = \sqrt{86}/5$. Let $(a_m)_{m>0}$ and $(b_m)_{m>0}$ be the sequences of rational numbers defined by $\alpha^m = a_m\alpha + b_m$, and let $q_m = a_m X + b_m \in \mathbb{Q}[X]$. It is easy to verify by induction on m that $a_1 = 1$, $b_1 = 0$,*

$$a_{m+1} = \frac{6}{5}a_m + b_m \text{ and } b_{m+1} = -a_m \text{ for } m > 0\,.$$

Solving the above recurrences yields

$$a_m = \frac{5}{8}\sqrt{-1}\left(\left(\frac{3-4\sqrt{-1}}{5}\right)^m - \left(\frac{3+4\sqrt{-1}}{5}\right)^m\right) = \frac{5}{4}\sin(m\theta)$$

and

$$b_m = -\frac{5}{8}\sqrt{-1}\left(\left(\frac{3-4\sqrt{-1}}{5}\right)^{m-1} - \left(\frac{3+4\sqrt{-1}}{5}\right)^{m-1}\right) = -\frac{5}{4}\sin((m-1)\theta)$$

where $\theta \in \mathbb{R}$ *is such that* $\alpha = e^{\theta\sqrt{-1}}$. *Therefore,*

$$|q_m| = \sqrt{a_m^2 + b_m^2} = \frac{5}{4}\sqrt{\sin(m\theta)^2 + \sin((m-1)\theta)^2} \le \frac{5}{4}\sqrt{2}.$$

The function $f(x) = \sqrt{\sin(x)^2 + \sin(x-\theta)^2}$ *is periodic and continuous on* $\mathbb{R}$ *so let* $\mu(f) = \min_{0\le x<2\pi}(f(x)) \ge 0$. *Since* $\sin(\theta) = 4/5$, θ *is not an integer multiple of* π, *so* $\mu(f) > 0$ *and we have*

$$0 < \frac{5}{4}\mu(f) \le \frac{5}{4}f(m\theta) = |q_m| \le \frac{5}{4}\sqrt{2},$$

which implies that there are real numbers $A < B$ *such that* $A \le \log|q_m| \le B$. *So for any function* $P(x,y,z)$ *continuous on* $(2, \log(\sqrt{86}/5), [A \dots B])$, *in particular any polynomial, choosing* $\beta = q_m(\alpha)$ *where* m *is larger than the maximum of* P *on the above domain yields an instance of the orbit problem whose solution is not bounded by* $P(n_\alpha, \log|f_\alpha|, \log|q|)$.

We now correct the upper bounds given in [7] for solving the orbit problem. If α, which is assumed not to be a root of unity, is an algebraic integer, then the bound of [7] is correct: they use the results of [4] and obtain the bound

$$m \le n_\alpha + 60n_\alpha^2 \ln(6n_\alpha)(\log(n_\alpha+1) + \log|q|). \qquad (1)$$

If α is not an algebraic integer, let then $B, D \in \mathbb{Z}$ be such that $B > 0$, $D > 0$, $B\alpha$ is an algebraic integer and $Dq \in \mathbb{Z}[X]$. B can be chosen to be the least common multiple of the denominators of the coefficients of f_α (but B is not always bounded by $|f_\alpha|$ as is wrongly stated in [7], see Example 2), and D can be chosen to be the least common multiple of the denominators of the coefficients of q. Since $\deg(q) < n_\alpha$, $B^{n_\alpha-1}Dq(\alpha)$ is an algebraic integer. However $B^{n_\alpha}q(\alpha)$ is not always an algebraic integer (as wrongly stated in [7]) since it can happen that $D > B$ as in Example 2. Using the unique factorization of ideals in number rings as in [7], we then obtain the bound

$$m \le (n_\alpha - 1)\log B + \log D \qquad (2)$$

on the solution of the orbit problem. Together, bounds (1) and (2) provide a complete solution to the orbit problem. This new bound is polynomial in n_α, $\log|q|$ and in the number of bits required to express the coefficients of f_α and q as exact rational numbers: let $M \in \mathbb{Z}$ be the maximum of the absolute values of the denominators of those coefficients. Since we can choose $B \le M^{n_\alpha}$ and $D \le M^{n_\alpha}$, it follows that the bound (2) is at most $n_\alpha^2 \log M$.

Example 2. *Using the same* α *as in Example 1, we see that* $B = 5$ *is the smallest positive integer such that* $B\alpha$ *is an algebraic integer, while* $|f_\alpha| = \sqrt{86}/5 < B$. *It follows from the expressions for* a_m *and* b_m *that* $5^{m-1}8q_m \in \mathbb{Z}[X]$. *Therefore, the bound given by (2) for the orbit problem* $\alpha^n = q_m(\alpha)$ *is*

$$n \le \log B + \log D = \log 5 + \log(5^{m-1}8) = m\log 5 + \log 8 \approx 2.3m + 3,$$

which is correct since the solution is $n = m$.

Bounds (1) and (2) provide a worst–case guaranteed algorithm but a smaller set of candidate solutions can be obtained for most practical instances of the orbit problem: whenever we can compute $A, B, C, D \in \mathbb{Q}$ such that $A < |\beta| < B$ and either $C < |\alpha| < D < 1$ or $1 < C < |\alpha| < D$, then the constraint $(C^n, D^n) \cap (A, B) \ne \emptyset$ yields a (generally small) number of possibilities for n. Finally we note that they are other algorithms for special cases of the orbit problem that can yield better bounds: Shank [11] and Abramov [2] both describe algorithms for the quadratic case. For $M \in \mathbb{Z}$ larger than the absolute values of the numerators and denominators of the coefficients of f_α and q, the number of arithmetic operations of the method of [2] is $O(\log\log M)$ and its bit–complexity is $O(\log M(\log\log M)^2 \log\log\log M)$.

4. THE POWER ROOT PROBLEM

We discuss in this section the variant of the orbit problem that arises when solving q–difference equations, as well as a more general bivariate version. When bounding the degree n of the polynomial solutions of a q–difference equation with coefficients in a polynomial ring $C[X]$, one obtains an indicial equation $p \in C[X]$ similar to the one used for differential or difference equations. But instead of having $p(n) = 0$ as in the classical cases, the condition on the degree is $p(q^n) = 0$ [3] where $q \in C$ is such that $\sigma X = qX$. This is an instance of the *power root problem*: given α algebraic over a field F and a polynomial p with coefficients in F, determine all the integers $n > 0$ such that $p(\alpha^n) = 0$. Assuming that we can determine all the roots in $F(\alpha)$ of polynomials with coefficients in F, which is always the case if we can factor elements of $F(\alpha)[X]$ into irreducibles, then it simply means solving the orbit problems $\alpha^n = \beta_i$, where $\beta_1, \dots, \beta_m \in F(\alpha)$ are the roots of p in $F(\alpha)$.

Example 3. *Consider the* q–*difference equation*

$$4(2x+1)y(2x) - 3\sqrt{2}y(x\sqrt{2}) + (x+1)y(x) = 0. \qquad (3)$$

If it has a solution $y \in \mathbb{Q}(\sqrt{2})(x)$ *with a denominator of the form* $x^m d(x)$ *where* $m \ge 0$ *and* $d(0) \ne 0$, *then* $\sqrt{2}^m$ *is a root of the indicial equation* $Z^2 - 3\sqrt{2}Z + 4 = 0$ *(see [3]). This is an instance of the power root problem. Since the indicial equation factors in* $\mathbb{Q}(\sqrt{2})[Z]$ *as*

$$Z^2 - 3\sqrt{2}Z + 4 = (Z - 2\sqrt{2})(Z - \sqrt{2})$$

we must solve the orbit problems $\alpha^m = \beta_1$ *and* $\alpha_m = \beta^2$ *where* $\alpha = \beta_1 = \sqrt{2}$ *and* $\beta_2 = 2\sqrt{2}$. *We have* $n_\alpha = 2$, $q = X$ *for* β_1 *and* $q = 2X$ *for* β_2, *so (1) yields the upper bounds* $m < 948$ *for* β_1 *and* $m < 1544$ *for* β_2. *In that particular case, since*

$$1 < \frac{7}{5} < |\alpha| < \frac{36}{25}, \frac{7}{5} < |\beta_1| < \frac{36}{25} \text{ and } \frac{14}{5} < |\beta_2| < \frac{72}{25}$$

we obtain the necessary conditions

$$\left(\frac{7}{5}\right)^m < \frac{36}{25} \text{ for } \beta_1 \quad \text{and} \quad \left(\frac{7}{5}\right)^m < \frac{72}{25} \text{ for } \beta_2\,.$$

Taking logarithms on both sides yields the improved bounds $m < 2$ for β_1 and $m < 4$ for β_2. Checking powers finally yields the solutions $\beta_1 = \alpha^1$ and $\beta_2 = \alpha^3$. It turns out that $1/x^3$ is indeed a rational solution of (3).

A more interesting variant arises when one computes dispersions through resultants in $C(n)[t]$ with respect to the automorphism σ that maps n to $n+1$ and t to at for some $a \in C(n)$. This yields an instance of the following bivariate power root problem: given α algebraic over a field F and a bivariate polynomial p with coefficients in F, determine all the integers $n > 0$ such that $p(n, \alpha^n) = 0$. To solve it, we need the following variant of Theorem 1 of [10].

LEMMA 2. *Let $p, q \in K[X]$ be polynomials over a field K, with $p \neq 0$, q monic and $n = \deg(q) > 0$. Let $Y_0, \dots, Y_{n-1}$ be indeterminates over K and write the remainder of*

$$p(Y_0 + Y_1 X + \cdots + Y_{n-1} X^{n-1})$$

by q in $K[Y_0, \dots, Y_{n-1}][X]$ as $V_{n-1}X^{n-1} + \cdots + V_1 X + V_0$ where $V_i \in K[Y_0, \dots, Y_{n-1}]$. If q is squarefree, then the system of equations

$$V_0(Y_0, \dots, Y_{n-1}) = \cdots = V_{n-1}(Y_0, \dots, Y_{n-1}) = 0 \qquad (4)$$

has at most $\deg(p)^{\deg(q)}$ solutions in the algebraic closure $\overline{K}$ of K.

PROOF. Because of differences between our lemma and Osipov's result [10], we present a modification of his proof adapted for our case. Since q is squarefree, let $\alpha_1, \dots, \alpha_n$ be its distinct roots in $\overline{K}$, and let $\beta_1, \dots, \beta_k$ be the distinct roots of p in $\overline{K}$, where $k \leq \deg(p)$. For any root $(y_0, \dots, y_{n-1}) \in \overline{K}^n$ of (4) and for $1 \leq i \leq n$ we have

$$\begin{aligned} p(y_0 + y_1\alpha_i + \cdots + y_{n-1}\alpha_i^{n-1}) \quad &= \\ V_0(y_0, \dots, y_{n-1}) \quad &+ \quad V_1(y_0, \dots, y_{n-1})\alpha_i \\ &+ \cdots + V_{n-1}(y_0, \dots, y_{n-1})\alpha_i^{n-1} = 0\,. \end{aligned}$$

Therefore, $y_0 + y_1\alpha_i + \cdots + y_{n-1}\alpha_i^{n-1} \in \{\beta_1, \dots, \beta_k\}$ for $1 \leq i \leq n$, so $(y_0, \dots, y_{n-1})$ is a solution of the linear system

$$\begin{pmatrix} 1 & \alpha_1 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \dots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_n & \dots & \alpha_n^{n-1} \end{pmatrix} \begin{pmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{n-1} \end{pmatrix} = \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{n-1} \end{pmatrix} \qquad (5)$$

where the right–hand side is in $\{\beta_1, \dots, \beta_k\}^n$. This means that there are at most k^n possible systems of the form (5), each having a unique solution in $\overline{K}^n$ since its matrix is an invertible Vandermonde. Therefore, the system (4) has at most k^n solutions in $\overline{K}$. $\square$

We also make use of the following classical result, whose proof is a simple exercise in high–school mathematics, for bounding the numerators and denominators of fractional zeros of polynomials with integer coefficients.

LEMMA 3. *Let $p = \sum_{i=a}^{b} p_i X^i$ be a polynomial over a unique factorization domain R with $a \leq b$. For any $q \neq 0$ in the quotient field of R,*

$$p(q) = 0 \implies p_b q \in R \text{ and } p_a q^{-1} \in R\,.$$

We now have the following algorithm for solving the bivariate power root problem over $\mathbb{Q}$.

THEOREM 3. *If $\alpha \in \overline{\mathbb{Q}}$ is not a root of unity, then for any $p \in \mathbb{Q}[Z, Y]$ with $\deg_Z(p) > 0$, the set of positive integer solutions of $p(n, \alpha^n) = 0$ is finite, and it is possible to compute an upper bound for that set.*

PROOF. Let $f_\alpha \in \mathbb{Q}[X]$ is the monic minimal polynomial for α over $\mathbb{Q}$, $n_\alpha = \deg(f_\alpha)$ and write the remainder of

$$p(Z, Y_0 + Y_1 X + \cdots + Y_{n_\alpha - 1} X^{n_\alpha - 1})$$

by f_α in $\mathbb{Q}[Z, Y_0, \dots, Y_{n_\alpha-1}][X]$ as $V_{n_\alpha-1}X^{n_\alpha-1} + \cdots + V_1 X + V_0$ where $V_i \in \mathbb{Q}[Z, Y_0, \dots, Y_{n_\alpha-1}]$. Let I be the ideal of $\mathbb{Q}[Z, Y_0, \dots, Y_{n_\alpha-1}]$ generated by $(V_0, \dots, V_{n_\alpha-1})$. Lemma 2 applied to p, $q = f_\alpha$ and $K = \mathbb{Q}(Z)$ implies that for $0 \leq i < n_\alpha$, $I \cap \mathbb{Q}[Z, Y_i]$ contains some polynomial U_i such that $\deg_{Y_i}(U_i) > 0$ (the system (4) would have infinitely many solutions otherwise). Furthermore, such U_i's can be computed using elimination techniques (*e.g.* resultants or Gröbner bases) and any $(z, y_0, y_1, \dots, y_{n_\alpha-1}) \in \overline{\mathbb{Q}}^{n+1}$ satisfying $p(z, y_0 + y_1\alpha + \dots y_{n_\alpha-1}\alpha^{n_\alpha-1}) = 0$ must also be a zero of I, so it is a solution of

$$U_0(z, y_0) = \cdots = U_{n_\alpha-1}(z, y_{n_\alpha-1}) = 0\,. \qquad (6)$$

Multiplying them by sufficiently large integers, we can assume that $U_i \in \mathbb{Z}[Z, Y_i]$ for each i, so write them as

$$U_i = W_{i,M_i} Y_i^{M_i} + \cdots + W_{i,m_i} Y_i^{m_i}$$

where $W_{ij} \in \mathbb{Z}[Z]$, $W_{i,M_i} \neq 0$ and $W_{i,m_i} \neq 0$. Let now

$$\mathcal{S} = \bigcup_{0 \leq i < n_\alpha} \{m \in \mathbb{Z} \text{ s.t. } W_{i,M_i}(m) = 0 \text{ or } W_{i,m_i}(m) = 0\}\,.$$

Since $\mathcal{S}$ is finite, the solutions $m \in \mathcal{S}$ of $p(m, \alpha^m) = 0$ can be found by exhaustive search. Let now $m > 0$ be a solution outside $\mathcal{S}$ of the bivariate power root problem, *i.e.* $m \notin \mathcal{S}$ and $p(m, \alpha^m) = 0$, and let $q = q_0 + q_1 X + \dots q_{n_\alpha-1} X^{n_\alpha-1}$ be the unique polynomial in $\mathbb{Q}[X]$ of degree at most $n_\alpha - 1$ such that $\alpha^m = q(\alpha)$. Then, $(m, q_0, q_1, \dots, q_{n_\alpha-1})$ is a solution of the system (6), which implies that each q_i is a root of

$$Q_i = W_{i,M_i}(m) Y^{M_i} + \cdots + W_{i,m_i}(m) Y^{m_i} \in \mathbb{Z}[Y]\,.$$

Lemma 3 then implies that $W_{i,M_i}(m) q_i$ and $W_{i,m_i}(m) q_i^{-1}$ are both in $\mathbb{Z}$. Suppose first that α is an algebraic integer. Since $m \notin \mathcal{S}$, $W_{i,m_i}(m) \neq 0$ must be a multiple of the numerator of q_i, which implies that $|q_i| \leq |W_{i,m_i}(m)|$, hence that

$$|q| \leq \sqrt{W_{0,m_0}^2(m) + \cdots + W_{n_\alpha-1,m_{n_\alpha-1}}^2(m)}\,.$$

By the results of Section 3, the inequality (1) must hold since $\alpha^m = q(m)$, so combining with the above bound on $|q|$ we obtain

$$m \leq c_1 + c_2 \log\left(W_{0,m_0}^2(m) + \cdots + W_{n_\alpha-1,m_{n_\alpha-1}}^2(m)\right) \quad (7)$$

where

$$c_1 = n_\alpha + 60n_\alpha^2 \ln(6n_\alpha)\log(n_\alpha+1) \text{ and } c_2 = 30n_\alpha^2 \ln(6n_\alpha)\,.$$

Suppose now that α is not an algebraic integer and let $B \in \mathbb{Z}$ be such that $B > 0$ and $B\alpha \in \mathbb{Z}$. Since $W_{i,M_i}(m) \neq 0$ must be a multiple of the denominator of q_i, $W(m)q \in \mathbb{Z}$ where $W = \mathrm{lcm}(W_{0,M_0}, \ldots, W_{n_\alpha-1,M_{n_\alpha-1}}) \in \mathbb{Z}[Z]$. By the results of Section 3, the inequality (2) must hold, which implies that

$$m \le (n_\alpha - 1)\log B + \log|W(m)|\,. \tag{8}$$

In all cases we have produced real constants $C_1 \ge 0$, $C_2 > 0$ and a polynomial $Q \in \mathbb{Z}[Z]$ such that for any solution $m > 0$ of $p(m, \alpha^m) = 0$, either $m \in \mathcal{S}$ or $m \le C_1 + C_2 \log|Q(m)|$. Since

$$\lim_{m\to+\infty} \frac{C_1 + C_2\log|Q(m)|}{m} = 0$$

and $\mathcal{S}$ is finite, this proves that there are finitely many solutions. We have a trivial bound for m when Q has degree 0, so suppose that $d = \deg(Q) > 0$, and write $Q = \sum_{i=0}^d Q_i Z^i$ with $Q_d \neq 0$. Then, $|Q(m)| \le \mu m^d$ for $m \ge 1$ where $\mu = \max(1, (d+1)\max_{0\le i\le d}|Q_i|)$, which implies that

$$1 \le m \le C_1 + C_2\log|Q(m)| \le B_1 + B_2 \log m$$

where $B_1 = C_1 + C_2 \log\mu \ge 0$ and $B_2 = C_2 d > 0$. Let g be the function mapping $x > 0$ to $x - (B_1 + B_2\log x)$. Since $dg/dx = 1 - B_2/x$ is positive for $x > B_2$, g is monotonically increasing for $x > B_2$. If $g(B_2) \ge 0$, then $x > B_1 + B_2 \log x$ for $x > B_2$, which implies that $m \le B_2$. Otherwise, since g is smooth and strictly increasing on $(B_2, +\infty)$, it has exactly one zero in $(B_2, +\infty)$, which can be approximated easily since dg/dx tends to 1 as x tends to infinity. The ceiling of any such approximation is an upper bound for m. □

5. REFERENCES

[1] S. A. Abramov. On the summation of rational functions. *USSR Computational Mathematics and Mathematical Physics*, 11:324–330, 1971.

[2] S. A. Abramov. Complexity of the solution of exponential equation and the orbit problem for second-order matrices. *Moscow University Computational Mathematics and Cybernetics*, 4:55–63, 1987.

[3] S. A. Abramov. Rational solutions of linear difference and q–difference equations with polynomial coefficients. *Programming and Computer Software*, 21:273–278, 1995.

[4] P. Blanksby and H. Montgomery. Algebraic integers near the unit circle. *Acta Arithmetica*, 18:355–369, 1971.

[5] M. Bronstein. On solutions of linear ordinary difference equations in their coefficient field. *Journal of Symbolic Computation*, 29(in press), 2000. Also available as INRIA Research Report RR–3797.

[6] M. Harrison. *Lectures on sequential machines.* Academic Press, Orlando, 1969.

[7] R. Kannan and R. Lipton. Polynomial–time algorithm for the orbit problem. *Journal of the ACM*, 33(4):808–821, October 1986.

[8] M. Karr. Summation in finite terms. *Journal of the ACM*, 28:305–350, Apr. 1981.

[9] M. Karr. Theory of Summation in Finite Terms. *J. Symbolic Computation*, 1(3):303–316, September 1985.

[10] N. Osipov. On the simplification of nested real radicals. *Programming and Computer Software*, 23:142–146, 1997.

[11] H. Shank. The rational case of a matrix problem of Harrison. *Discrete Mathematics*, 28:207–212, 1979.