Quantifier rank for parity of embedded finite models

Hervé Fournier

Laboratoire de l'Informatique du Parallélisme
École Normale Supérieure de Lyon
46, allée d'Italie, 69364 Lyon Cedex 07, France

Abstract

We are interested in the quantifier rank necessary to express the parity of an embedded set of cardinal smaller than a given bound. We consider several embedding structures such as the reals with addition and order, or the field of complex numbers. We provide both lower and upper bounds. We obtain from these results some bounds on the quantifier rank needed to express the connectivity of an embedded graph, when a bound on its number of vertices is given.

Key words: Parity, Connnectivity, Reachability, Ehrenfeucht-Fraïssé games, Constraint databases

1 Introduction

There are numerous works about the expressiveness obtained by embedding a finite structure into an infinite one M. These studies have been carried out because of their fundamental role in the constraint database model. Among these results, the generic collapse results are of great importance. They state that embedding a finite model into some infinite structures does not help to express a large class of queries, called generic. These results hold for structures M having some good model-theoretic properties. The strongest result deals with structures without the independence property (1). One of these generic queries is parity, which asks if the cardinal of a finite set $\mathcal{I}$ is even. It follows from a special case of general collapse theorems, for some structures M, that there is no first-order sentence defining parity (2; 1; 3; 5) – for more references see the book (11). However, when restricting to the case where $|\mathcal{I}|$ is smaller

Email address: herve.fournier@ens-lyon.fr (Hervé Fournier).

Preprint submitted to Elsevier Science *5 March 2002*

than a given bound, such a formula obviously exists. In this paper we give upper and lower bounds on the quantifier rank of such formulas. Can we do better than in the case where the finite set stands alone?

Main results are summarized in figure 1. The first column presents the quantifier rank needed for a formula to express that a given set of cardinal bounded by n is even, when this set is embedded into a structure to be read on the left. The last column gives some results concerning connectivity of an embedded graph, where the bound n is on the number of vertices. The first two lines are well-known bounds from finite model theory. Comparing them to the next lines allows oneself to understand how addition and product can (or not) improve these bounds. For example, concerning parity, adding the addition to $(\mathbb{C}, =)$ makes the quantifier rank decrease exponentially (from n to $\lceil \log n \rceil + 1$), but adding further the product allows no gain at all. On the reals with order, the addition allows to decrease the quantifier rank for parity from $\lfloor \log n \rfloor + 1$ to $\Theta(\sqrt{\log n})$. It would be interesting to have precise bounds when the embedding structure is a real closed field, for example $(\mathbb{R}, +, -, \times, \leqslant)$ – see question 1 about this. At last, it was pointed out in (8) that parity and connectivity are expressible over $(\mathbb{Q}, +, -, \times, =)$: this comes from the definability of the integers over this structure (12; 7) (allowing the power of arithmetic).

Remark. It is natural to consider a graph embedded into M^k instead of M – think of geographic databases with $M = \mathbb{R}$. In this case, one consider the structure M equipped with the usual operations, and multiplies by k the arities of the vertices and edges of the graph. The results proved in this paper still hold in this case : lower bounds remain unchanged, while upper bounds have to be multiplied by k.

	Parity	Connectivity
$(\mathbb{Q}, =)$, $(\mathbb{R}, =)$, $(\mathbb{C}, =)$	n	$\Theta(\log n)$
$(\mathbb{Q}, <)$, $(\mathbb{R}, <)$	$\lfloor \log n \rfloor + 1$	$\Theta(\log n)$
$(\mathbb{Q}, +, -, =)$, $(\mathbb{R}, +, -, =)$, $(\mathbb{C}, +, -, =)$	$\lceil \log n \rceil + 1$	$\Theta(\sqrt{\log n})$
$(\overline{\mathbb{Q}}, +, -, \times, =)$, $(\mathbb{C}, +, -, \times, =)$	$\lceil \log n \rceil + 1$	$\Theta(\sqrt{\log n})$
$(\mathbb{Q}, +, -, <)$, $(\mathbb{R}, +, -, <)$	$\Theta(\sqrt{\log n})$	$\Theta(\sqrt{\log n})$
$(\mathbb{R}, +, -, \times, <)$	$\Omega(\log \log n)$ $O(\sqrt{\log n})$	$\Omega(\log \log n)$ $O(\sqrt{\log n})$
$(\mathbb{Q}, +, -, \times, =)$, $(\mathbb{Q}, +, -, \times, <)$	$\Theta(1)$	$\Theta(1)$

Fig. 1. Quantifier rank for parity and connectivity

Organization. The question we deal with is formally defined in section 2, and first remarks are made there. In section 3, we show some upper and lower bounds for parity in zero characteristic algebraically closed fields and $\mathbb{Q}$-vector spaces (for instance, the reals with addition and equality). Section 4 deals with ordered $\mathbb{Q}$-vector spaces (for instance, the reals with addition and order). We derive from these results some bounds on connectivity of embedded graphs in section 5. In the last section, we relate these bounds with the notion of active-natural collapse.

2 Notations and first remarks

We are interested in the following problem. We embed a finite set $\mathcal{I}$ in either an algebraically closed field or an ordered $\mathbb{Q}$-vector space: we shall call M this structure. Thus, besides the signature of M, we consider a new predicate I that is interpreted as $\mathcal{I}$. We shall be interested in the query Even, asking if $|\mathcal{I}|$ is even, and Card_m which asks if $|\mathcal{I}| \geqslant m$. For a query Q, $\mathrm{QR}_M(Q, n)$ denotes the smallest possible quantifier rank of a first-order formula expressing the query Q when it is known that $|\mathcal{I}| \leqslant n$. Our aim is to find some bounds on $\mathrm{QR}_M(\mathrm{Even}, n)$. We recall that the quantifier rank $\mathrm{qr}(\phi)$ of a formula ϕ is defined by induction on its structure. If ϕ is an atomic formula, $\mathrm{qr}(\phi) = 0$. Otherwise $\mathrm{qr}(\phi \vee \psi) = \mathrm{qr}(\phi \wedge \psi) = \max(\mathrm{qr}(\phi), \mathrm{qr}(\psi))$ and $\mathrm{qr}(\exists x \phi) = \mathrm{qr}(\forall x \phi) = 1 + \mathrm{qr}(\phi)$. Now let us make the following remark.

Lemma 1 *If two structures M and M' are elementarily equivalent, then for all n_0, we have* $\mathrm{QR}_M(\mathrm{Even}, n_0) = \mathrm{QR}_{M'}(\mathrm{Even}, n_0)$.

Proof. Indeed, let n_0 be fixed and suppose we have a first-order formula ϕ such that if $|\mathcal{I}| \leqslant n_0$, $(M, \mathcal{I}) \models \phi$ if and only if $|\mathcal{I}|$ is even. Let M' be a structure elementarily equivalent to M, and $n \leqslant n_0$. Let $\tilde{\phi}(x_1, \ldots, x_n)$ be the formula ϕ where $I(x)$ is replaced with $\bigvee_{i=1}^n x = x_i$, and $\psi_n = \forall x_1, \ldots, x_n \ \bigwedge_{i<j} x_i \neq x_j \to \tilde{\phi}(x_1, \ldots, x_n)$. If n is even, $M' \models \psi_n$ since $M \models \psi_n$; and if n is odd, $M' \models \neg\psi_n$. Thus if $\mathcal{I}' \subset M'$ with $|\mathcal{I}'| \leqslant n_0$, $(M', \mathcal{I}') \models \phi$ if $|\mathcal{I}'|$ is even and $(M', \mathcal{I}') \models \neg\phi$ if $|\mathcal{I}'|$ is odd. Hence $\mathrm{QR}_{M'}(\mathrm{Even}, n_0) \leqslant \mathrm{QR}_M(\mathrm{Even}, n_0)$ and by symmetry $\mathrm{QR}_{M'}(\mathrm{Even}, n_0) = \mathrm{QR}_M(\mathrm{Even}, n_0)$. □

Of course the previous remark also applies to the queries Card_m. This justifies the notation $\mathrm{QR}_T(\mathrm{Even}, n)$ and $\mathrm{QR}_T(\mathrm{Card}_m, n)$ for a complete theory T. Let us introduce some notations for the theories we shall be interested in, and give some examples of models of these theories.

- zero characteristic algebraically closed field: ACF_0
  $(\overline{\mathbb{Q}}, +, -, \times, =)$, $(\mathbb{C}, +, -, \times, =)$.
- $\mathbb{Q}$-vector space: $\mathbb{Q}vs$
  $(\mathbb{Q}, +, -, =)$, $(\mathbb{R}, +, -, =)$, $(\mathbb{C}, +, -, =)$.

- ordered $\mathbb{Q}$-vector space: $\mathcal{O}vs$
  $(\mathbb{Q}, +, -, <)$, $(\mathbb{R}, +, -, <)$.

Our main tool will be the back-and-forth games defined by Ehrenfeucht and Fraïssé (6; 10; 9). Consider two $\mathcal{L}$-structures M and N. A game of length n between the two structures M and N proceeds as follows. At the i-th step, the first player chooses a point either in M or N; then the second player must choose an element in the other structure. Let us call a_i the point chosen in M and b_i the point chosen in N. After n moves, the game ends, and the second player wins iff the same atomic formulas are true in the structures $(M, a_1, \ldots, a_n)$ and $(N, b_1, \ldots, b_n)$. The second player is said to have a strategy to win the game of length n between M and N if he can win no matter what the first player plays. We shall use the following fundamental property of Ehrenfeucht-Fraïssé games.

Fact 1 *If the second player has a strategy to win back-and-forth games of length n between two structures M and N, then the same formulas of quantifier rank at most n are true in M and N.*

We shall use this fact to establish lower bounds. For example, if we want to show that $\mathrm{QR}_T(\mathrm{Even}, n) > B$ for a given complete theory T of signature $\mathcal{L}$, we may proceed as follows. We choose two models of M and N of T and two finite sets $\mathcal{I} \subset M$ and $\mathcal{J} \subset N$ of cardinal at most n, with $|\mathcal{I}|$ odd and $|\mathcal{J}|$ even. Now if we show that the second player has a strategy to win the game of length B between the two $\mathcal{L} \cup \{I\}$-structures $(M, \mathcal{I})$ and $(N, \mathcal{J})$, then of course no first-order formula over $\mathcal{L} \cup \{I\}$ of quantifier rank at most B can express the restricted parity we are looking for.

Now let us examine some bounds on the quantifier rank when the finite structure stands alone. When no order is available, we shall write $\mathrm{QR}_=(\mathrm{Even}, n)$ the minimal quantifier rank of a first-order formula expressing parity. In the same way, we write it $\mathrm{QR}_<(\mathrm{Even}, n)$ when the universe is totally ordered. By some usual back-and-forth games (6; 10), we have $\mathrm{QR}_=(\mathrm{Even}, n) = n$ and $\mathrm{QR}_<(\mathrm{Even}, n) = \lfloor \log n \rfloor + 1$.

3 Parity on unordered structures

In this section, we show results concerning parity in 0-characteristic algebraically closed fields and $\mathbb{Q}$-vector spaces. Because a 0-characteristic algebraically closed field is a $\mathbb{Q}$-vector space, $\mathrm{QR}_{ACF_0}(\mathrm{Even}, n) \leqslant \mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Even}, n)$. Thus this section is divided in two parts: on one hand we show a lower bound on $\mathrm{QR}_{ACF_0}(\mathrm{Even}, n)$, on the other hand we show an upper bound on

$\mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Even}, n)$.

3.1 Lower bound in an algebraically closed field

The algebraic closure of $A \subseteq \mathbb{C}$ will be denoted $\overline{A}$. In this section, we shall prove the following lower bound.

Theorem 1 $\mathrm{QR}_{ACF_0}(\mathrm{Even}, n) \geqslant \lceil \log n \rceil + 1$.

Proof. Thanks to Lemma 1, it is enough to show this lower bound in a given algebraically closed field of characteristic 0. We shall work in the field of the complex numbers $\mathbb{C}$. Let $\mathcal{M}$ be a set of 2^{n-1} algebraically independent elements of $\mathbb{C}$, and $\mathcal{N}$ a set of $2^{n-1}+1$ algebraically independent elements of $\mathbb{C}$. We are going to prove that the second player can win the back-and-forth game of length n between $(\mathbb{C}, \mathcal{M})$ and $(\mathbb{C}, \mathcal{N})$.

At the beginning, let $E_n = F_n = \overline{\mathbb{Q}}$ and $\varphi = \mathrm{Id}_{\overline{\mathbb{Q}}}$. In the following, φ is a partial mapping from the first structure $(\mathbb{C}, \mathcal{M})$ to the second one $(\mathbb{C}, \mathcal{N})$. Its domain (resp. image) contains the points chosen by the two players in the first (resp. second) structure. At each step we shall extend φ so that its domain and image contain the points newly chosen. When it remains j steps to play, E_j denotes the field where φ is defined and $F_j = \varphi(E_j)$. At each step, φ is an isomorphism of algebraically closed field expanded with a unary predicate from E_j onto F_j: this means that for all x in E_j, $x \in \mathcal{M}$ iff $\varphi(x) \in \mathcal{N}$. We also maintain the following property $\mathcal{P}_j$.

First $|\mathcal{M} \setminus E_j|, |\mathcal{N} \setminus F_j| \geqslant 2^{j-1}$. *Moreover, if there exists* $a \in \mathcal{M} \setminus E_j$ *and* $A \subset \mathcal{M} \setminus (E_j \cup \{a\})$ *such that* $a \in \overline{E_j \cup A}$, *then* $|A| \geqslant 2^{j-1}$. *And the corresponding property in* $(\mathbb{C}, \mathcal{N})$*: if there exists* $a \in \mathcal{N} \setminus F_j$ *and* $A \subset \mathcal{N} \setminus (F_j \cup \{a\})$ *such that* $a \in \overline{F_j \cup A}$, *then* $|A| \geqslant 2^{j-1}$.

First let us check that $\mathcal{P}_n$ is satisfied. We have $|\mathcal{M} \setminus E_n| = |\mathcal{M}| \geqslant 2^{n-1}$. Moreover, there is no $a \in \mathcal{M}$ with $a \in \overline{\mathbb{Q} \cup A}$ such that $A \subseteq \mathcal{M}$ and $a \notin A$ because elements of $\mathcal{M}$ are algebraically independent over $\mathbb{Q}$. And the same is true in $(\mathbb{C}, \mathcal{N})$.

Let us suppose that $n-j-1$ steps have been played. The isomorphism φ is defined on E_{j+1} and it remains $j+1$ steps to do. Property $\mathcal{P}_{j+1}$ is satisfied by induction hypothesis. By symmetry, we can assume that the point is chosen in $(\mathbb{C}, \mathcal{M})$. Let v be this point. We can also assume $v \notin E_{j+1}$. There are two cases. First case: $v \in \overline{E_{j+1} \cup \{a_1, \ldots, a_r\}}$ with $a_i \in \mathcal{M} \setminus E_{j+1}$ distinct and $r \leqslant 2^{j-1}$. Then we choose some distinct elements $b_1, \ldots, b_r$ in $\mathcal{N} \setminus F_{j+1}$ and we define $\varphi(a_i) = b_i$. Thus $E_j = \overline{E_{j+1} \cup \{a_1, \ldots, a_r\}}$. Let $F_j = \overline{\varphi(E_j)}$ and let us extend φ to an isomorphism of fields from E_j onto F_j. Let us show that $\mathcal{P}_j$ is satisfied. If there exists $d \in \mathcal{M} \setminus E_j$, with $d \in \overline{E_j \cup \{c_1, \ldots, c_l\}}$, $c_i \in \mathcal{M} \setminus (E_j \cup \{d\})$ and $l \leqslant 2^{j-1} - 1$, then $d \in \overline{E_{j+1} \cup \{a_1, \ldots, a_r, c_1, \ldots, c_l\}}$. But $r + l \leqslant 2^{j-1} + 2^{j-1} - 1 = 2^j - 1$. Therefore we should have $d \in E_{j+1}$ by property $\mathcal{P}_{j+1}$, this is absurd. We have the same property in $(\mathbb{C}, \mathcal{N})$. Moreover, $|\mathcal{M} \setminus E_j|, |\mathcal{N} \setminus F_j| \geqslant 2^j - 2^{j-1} = 2^{j-1}$ so $\mathcal{P}_j$ is satisfied. Exactly in the same way, we show that there are no other points from $\mathcal{M} \setminus E_{j+1}$ in E_j besides the a_i: if $d \in (\mathcal{M} \cap E_j) \setminus (E_{j+1} \cup \{a_1, \ldots, a_r\})$, then $d \in \overline{E_{j+1} \cup \{a_1, \ldots, a_r\}}$ and we conclude with $\mathcal{P}_{j+1}$. This also holds in $(\mathbb{C}, \mathcal{N})$, and it shows that φ is an isomorphism. Second case: let $f \notin \overline{F_{j+1} \cup \mathcal{N}}$. Let $\varphi(v) = f$. We set $E_j = \overline{E_{j+1} \cup \{v\}}$. Let $F_j = \overline{\varphi(E_j)}$ and let us extend φ to an isomorphism of fields from E_j onto F_j. Let us show that $\mathcal{P}_j$ is satisfied. Let $a \in \mathcal{M} \setminus E_j$ such that $a \in \overline{E_j \cup A}$ for $A \subseteq \mathcal{M} \setminus (E_j \cup \{a\})$ with $|A| < 2^{j-1}$. Thus $a \in \overline{E_{j+1} \cup \{v\} \cup A}$. This shows $v \in \overline{E_{j+1} \cup \{a\} \cup A}$, because $a \in \overline{E_{j+1} \cup A}$ is impossible by $\mathcal{P}_{j+1}$. But we should be in the first case since $|A \cup \{a\}| \leqslant 2^{j-1}$. This also holds in $(\mathbb{C}, \mathcal{N})$ by the choice of f. Moreover, there is no point of $\mathcal{M}$ in $E_j \setminus E_{j+1}$ because if $a \in \mathcal{M} \cap E_j \setminus E_{j+1}$, then $a \in \overline{E_{j+1} \cup \{v\}}$ and as $a \notin E_{j+1}$ we would have $v \in \overline{E_{j+1} \cup \{a\}}$ which is absurd. This also holds in $(L, \mathcal{N})$ thanks to the choice of f, thus φ remains an isomorphism. Moreover, $|\mathcal{M} \setminus E_j| = |\mathcal{M} \setminus E_{j+1}| \geqslant 2^{j-1}$ which ends to show $\mathcal{P}_j$. This ends the back-and-forth game. Thus we have shown $\mathrm{QR}_{ACF_0}(\mathrm{Even}, 2^{n-1} + 1) > n$. As $\mathrm{QR}_{ACF_0}(\mathrm{Even}, \cdot)$ is an increasing function, we obtain $\mathrm{QR}_{ACF_0}(\mathrm{Even}, n) \geqslant \lceil \log n \rceil + 1$. □

3.2 Upper bound in a $\mathbb{Q}$-vector space

The proof will proceed in three steps.

- First we show that it is possible to express $|\mathcal{I}| \geqslant m$ with a formula of quantifier rank $\lceil \log m \rceil + 2$ in the special case where the elements of $\mathcal{I}$ are known to be linearly independent over $\mathbb{Q}$.
- Then we generalize this bound to the general case.
- At last, we show how to decrease the quantifier rank of these formulas by 1.

We need to build a family of formulas $S_{(\alpha_1, \ldots, \alpha_p)}(x)$ for $p \geqslant 1$ and $(\alpha_1, \ldots, \alpha_p) \in \mathbb{N}^p$. These formulas should satisfy the following:

$$(M, \mathcal{I}) \models S_{(\alpha_1, \ldots, \alpha_p)}(x) \iff \exists x_1, \ldots, x_p \in \mathcal{I}\ x = \sum_{i=1}^{p} \alpha_i x_i \ .$$

Moreover, we shall design such formulas with small quantifier rank. We define them as follows: we take $S_{(1)}(x) := I(x)$, and $S_{(\alpha_1)}(x) := \exists y I(y) \wedge x = y + \ldots + y$

(where y is being added to itself α_1 times) for $\alpha_1 \neq 1$. At last, we define

$$S_{(\alpha_1,\ldots,\alpha_p)}(x) := \exists y\ S_{(\alpha_1,\ldots,\alpha_{\lfloor p/2 \rfloor})}(y) \wedge S_{(\alpha_{\lceil p/2 \rceil},\ldots,\alpha_p)}(x-y)\ .$$

One can check that the quantifier rank of $S_{(\alpha_1,\ldots,\alpha_p)}(x)$ is bounded above by $\lceil \log p \rceil + 1$.

Proposition 1 *In a $\mathbb{Q}$-vector space, if we restrict ourselves to the case where the elements of $\mathcal{I}$ are linearly independent over $\mathbb{Q}$, we can express that $|\mathcal{I}| \geqslant m$ with a formula of quantifier rank $\lceil \log m \rceil + 2$.*

Proof. Let $\bar{\alpha}_m$ be $(1,1,\ldots,1) \in \mathbb{N}^m$ and $\bar{\beta}_m$ be $(2,1,1,\ldots,1) \in \mathbb{N}^{m-1}$. Let us define $F_m = \exists x\ S_{\bar{\alpha}_m}(x) \wedge \neg S_{\bar{\beta}_m}(x)$. Note that $\mathrm{qr}(F_m) \leqslant \lceil \log m \rceil + 2$. We claim that F_m expresses $|\mathcal{I}| \geqslant m$. Indeed if F_m is true, this means that there exists x which is a sum of m *different* elements of $\mathcal{I}$: these elements must be different because the second part of F_m ensures that x is not a linear combination of $m-1$ elements of $\mathcal{I}$ with coefficients $(2,1,\ldots,1)$. Conversely, if $|\mathcal{I}| \geqslant m$, take s to be the sum of m different elements of $\mathcal{I}$. The formula $S_{\bar{\alpha}_m}(s) \wedge \neg S_{\bar{\beta}_m}(s)$ will be true because the elements of $\mathcal{I}$ are linearly independent, thus F_m will be true. □

Proposition 2 $\mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Card}_m, n) \leqslant \lceil \log m \rceil + 2$.

Proof. We shall work in $\mathbb{Q}$, and assume $|\mathcal{I}| \leqslant n$. Let us notice that if the formula described in the previous proof is true, then $|\mathcal{I}| \geqslant m$. And if it is false, then we don't know – because we do not have the hypothesis of linear independence anymore. To get rid of this hypothesis, the trick is to weigh the sum in the previous proof by some integer coefficients.

Let us notice that $S_{\bar{\alpha}}(x)$ is equivalent to $S_{\bar{\gamma}}(x)$ where $\bar{\gamma}$ is obtained from $\bar{\alpha}$ by permuting some elements. That is why we consider in the following only non-decreasing tuples. For a tuple $\bar{\alpha} = (\alpha_1,\ldots,\alpha_p) \in \mathbb{N}^p$, let us define $s(\bar{\alpha})$ to be the set of increasing tuples of $\mathbb{N}^{p-1}$ obtained by replacing in $\bar{\alpha}$ any two elements α_i and α_j by their sum. For example, $s((1,4,7)) = \{(5,7),(4,8),(1,11)\}$ and $s((1,2,2,3)) = \{(2,3,3),(2,2,4),(1,3,4),(1,2,5)\}$. Let us define the following formula

$$J_{\bar{\alpha}} = \exists x S_{\bar{\alpha}}(x) \wedge \bigwedge_{\bar{\beta} \in s(\bar{\alpha})} \neg S_{\bar{\beta}}(x)\ .$$

For the reason mentioned above, if $J_{(\alpha_1,\ldots,\alpha_p)}$ is true then $|\mathcal{I}| \geqslant p$.

We are ready to build a formula expressing $|\mathcal{I}| \geqslant m$, under the hypothesis $|\mathcal{I}| \leqslant n$. Let $N = m!n^{m-1}$. Let $\mathcal{A}$ be a set of $Nm+1$ elements of $\mathbb{N}^m$ in general position: by this we mean that no $m+1$ of these elements lie on a same hyperplane of $\mathbb{R}^m$. We claim that $H_m := \bigvee_{\bar{\alpha} \in \mathcal{A}} J_{\bar{\alpha}}$ is true if and only if $|\mathcal{I}| \geqslant m$.

Proof of the claim: if H_m is true, then for a $\bar{\alpha} \in \mathcal{A}$ the formula $J_{\bar{\alpha}}$ is true and this implies that $|\mathcal{I}| \geqslant m$. For the converse, let us suppose $|\mathcal{I}| \geqslant m$. Then $\mathcal{I} = \{x_1, \ldots, x_l\}$, where the x_i are all different, and $m \leqslant l \leqslant n$. Let us consider all the equations

$$\sum_{i=1}^{m} A_i x_i = \sum_{i=1}^{m-2} A_{\sigma(i)} x_{t(i)} + (A_{\sigma(m-1)} + A_{\sigma(m)}) x_{t(m-1)}$$

where σ runs over all the permutations of $\{1, \ldots, m\}$ and t runs over all the mappings from $\{1, \ldots, m-1\}$ to $\{1, \ldots, l\}$. These equations define a family $\mathcal{H}_{\bar{x}}$ of hyperplanes of $\mathbb{R}^m$ in $A_1, \ldots, A_m$ parameterized by $(x_1, \ldots, x_l)$. First these are all *true* hyperplanes (no equation is of the form "$0 = 0$" or "$0 = 1$"). Let us also remark that $|\mathcal{H}_{\bar{x}}| \leqslant N$. On each hyperplane of $\mathcal{H}_{\bar{x}}$ there are at most m elements of $\mathcal{A}$ since they are in general position. As $|\mathcal{A}| > |\mathcal{H}_{\bar{x}}|m$, there must be at least one $\bar{\alpha} \in \mathcal{A}$ which is not on any hyperplane of $\mathcal{H}_{\bar{x}}$. For such a $\bar{\alpha}$ the formula $J_{\bar{\alpha}}$ is true (take $x = \sum_{i=1}^{m} \alpha_i x_i$ for the first existential quantifier). Thus H_m is true, and the claim is proved. □

Theorem 2 $\mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Card}_m, n) \leqslant \lceil \log m \rceil + 1$.

Proof. Once again we assume $|\mathcal{I}| \leqslant n$. We just have to use one quantifier rank less than in the previous proposition. We can extend the definition of $S_{(\alpha_1,\ldots,\alpha_m)}(x)$ to the case where the α_i are rational. We define $S_{(1/q)}(x) := I(qx)$, where qx is $x + \ldots + x$ (q times). We also define $S_{(p/q)}(x) := \exists y\ I(qy) \wedge x = py$ when $p \neq 1$. As previously, we take $S_{(\alpha_1,\ldots,\alpha_m)}(x) := \exists y\ S_{(\alpha_1,\ldots,\alpha_{\lfloor m/2 \rfloor})}(y) \wedge S_{(\alpha_{\lceil m/2 \rceil},\ldots,\alpha_m)}(x - y)$. Now, we take $\mathcal{A}$ to be a set of $Nm + 1$ elements of $(1/(\mathbb{N} \setminus \{0\}))^m$ in general position. For $\bar{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathcal{A}$, what is the quantifier rank of $J_{\bar{\alpha}}$? The quantifier rank of $S_{\bar{\alpha}}(x)$ is $\lceil \log m \rceil$. Moreover, we claim that for each $\bar{\beta} \in s(\bar{\alpha})$, we can permute some β_i in order to obtain $\tilde{\beta}$ such that $\mathrm{qr}(S_{\tilde{\beta}}(x)) = \lceil \log m \rceil$ (of course the formula $S_{\tilde{\beta}}(x)$ is equivalent to $S_{\bar{\beta}}(x)$). Notice that, in $\bar{\beta}$, all coefficients have numerator 1 except maybe the one which is of the form $\alpha_i + \alpha_j$. Two cases may happen. If $m - 1$ is a power of 2, then $\mathrm{qr}(S_{\bar{\beta}}(x)) = \lceil \log(m-1) \rceil + 1 = \lceil \log m \rceil$. Let us assume now that $m - 1$ is not a power of 2. The formula $S_{\bar{\beta}}(x)$ has the shape of a binary tree with $m - 1$ leaves, its height being $\lceil \log(m-1) \rceil$. To each leaf is attached a formula $S_{(\beta_i)}(\cdot)$. As $m - 1$ is not a power of two, at least one leaf has depth $\lceil \log(m-1) \rceil - 1$. The formula $S_{(p/q)}(\cdot)$ with $p \neq 1$ must be attached to such a leaf: this can be achieved by swapping this coefficient p/q with another one. This ensures that $\mathrm{qr}(S_{\tilde{\beta}}(x)) = \lceil \log(m-1) \rceil$. Therefore we can build a formula equivalent to $J_{(\alpha_1,\ldots,\alpha_m)}$ with quantifier rank $\lceil \log m \rceil + 1$. □

Corollary 1 $\mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Even}, n) = \mathrm{QR}_{ACF_0}(\mathrm{Even}, n) = \lceil \log n \rceil + 1$.

Proof. Let n be fixed. By Theorem 2, for any $m \leqslant n$, there is a formula F_m expressing $|\mathcal{I}| \geqslant m$, with $\mathrm{qr}(F_m) = \lceil \log m \rceil + 1 \leqslant \lceil \log n \rceil + 1$. Of course $F_m \wedge \neg F_{m+1}$ expresses that $|\mathcal{I}| = m$. Now if we know that $|\mathcal{I}| \leqslant n$, $|\mathcal{I}|$ is even if and only if $\bigvee_{2k\leqslant n} |\mathcal{I}| = 2k$. Observe that $|\mathcal{I}| \geqslant n$ is equivalent to $|\mathcal{I}| = n$ since we know that $|\mathcal{I}| \leqslant n$. Thus our formula expressing parity will be $\bigvee_{2k\leqslant n} F_{2k} \wedge \neg F_{2k+1}$ when n is odd, and $\bigvee_{2k<n}(F_{2k} \wedge \neg F_{2k+1}) \vee F_n$ when n is even. That allows to obtain the desired upper bound. The lower bound was established in Theorem 1. □

Remark 1 *The proof of Theorem 1 goes unchanged when working over an algebraically closed field of positive characteristic. Concerning the upper bound of Theorem 2, it is even easier to prove such a result for vector spaces over a finite field. Therefore, we have*

$$\mathrm{QR}(\mathrm{Even}, n) = \lceil \log n \rceil + 1$$

over any algebraically closed field or infinite vector space.

4 Parity on ordered structures

We recall that $\mathcal{O}vs$ is the notation used for the theory of $\mathbb{Q}$-ordered vector spaces. We first show a lower bound. We define N_p the following way:

$$\begin{cases} N_0 = & 1 \\ N_{p+1} = & (2^p + 1)N_p \end{cases}$$

For any finite ordered set S, we define d_∞ from S^2 to $\mathbb{Z}$ as follows. For $x \leqslant y$ in S, we define $d_\infty(x, y) = |\{z \in S,\ x < z \leqslant y\}|$. Then, for $j \in \mathbb{N}$, we define $d_j(x, y) = d_\infty(x, y)$ if $d_\infty(x, y) < N_j$, $d_j(x, y) = \infty$ otherwise. At last, we take $d_\infty(y, x) = -d_\infty(x, y)$ and $d_j(y, x) = -d_j(x, y)$. First we need a simple remark.

Lemma 2 *We consider a modified version of the back-and-forth game between two finite ordered sets A and B where it is possible to choose up to 2^j elements at once, on the same side, when it remains j moves to play. If $|A|, |B| \geqslant N_{n+1}$, then the second player has a strategy to win this modified game of length n between A and B.*

Proof. Let $|A|, |B| \geqslant N_{n+1}$. We show how to play a game of length n. Before the game begins, we define our partial isomorphism α to send the minimum and maximum of A onto the minimum and maximum of B. We can assume that it remains j moves to play. Let us call $D \subseteq A$ the set where α is defined. By induction hypothesis, we assume that $d_{j+1}(a, a') = d_{j+1}(\alpha(a), \alpha(a'))$ for $a, a' \in D$. We proceed as in the case of back-and-forth games between two finite linear orderings – see (6; 10) – except that we can take 2^j elements in one move. We shall handle all at once the elements $a_1 < a_2 < \ldots < a_k$ lying in an interval

$]c,d[$ with $c,d \in D$, $]c,d[\cap D=\emptyset$. First case: $d_{j+1}(c,d)<\infty$. By induction, $d_{j+1}(c,d)=d_{j+1}(\alpha(c),\alpha(d))$ and we chose the $\alpha(a_i)$ in the obvious way. Second case: $d_{j+1}(c,d)=\infty$. Let $a_0=c$ and $a_{k+1}=d$. We successively choose $\alpha(a_l)$ for $l=1,2,\ldots,s$ such that $d_j(a_l,a_{l+1})=d_j(\alpha(a_l),\alpha(a_{l+1}))$, where s is the smallest subscript such that $d_j(a_s,a_{s+1})=\infty$. We proceed in the same way for $l=k,k-1,\ldots,t$ where t is the larger subscript such that $d_j(a_{t-1},a_t)=\infty$. If the images of all the a_i for $1\leqslant i\leqslant k$ have not yet been determined, then we successively choose the images of a_l for $l=s+1,\ldots,t-1$: we choose $\alpha(a_l)$ such that $d_\infty(\alpha(a_{l-1}),\alpha(a_l))=\min\{N_j,d_j(a_{l-1},a_l)\}$. Let us show we have enough points from B in $]\alpha(a_0),\alpha(a_{k+1})[$. As $d_{j+1}(\alpha(a_0),\alpha(a_{k+1}))=\infty$ by induction, we have $d_\infty(\alpha(a_0),\alpha(a_{k+1}))\geqslant N_{j+1}$. Taking into account that $k\leqslant 2^j$ and $N_{j+1}=(2^j+1)N_j$, there are indeed enough points to proceed this way. □

Theorem 3 $\mathrm{QR}_{\mathcal{O}vs}(\mathrm{Even},n)=\Omega(\sqrt{\log n})$.

Proof. We recall that all ordered $\mathbb{Q}$-vector spaces have the same first-order theory. By Lemma 1, this justifies the notation $\mathrm{QR}_{\mathcal{O}vs}(\mathrm{Even},n)$. In order to prove the result, we shall choose two ordered vector spaces expanded with a unary predicate $(V,\mathcal{M})$ and $(W,\mathcal{N})$ such that $|\mathcal{M}|=N_{n+1}$ and $|\mathcal{N}|=N_{n+1}+1$. We shall prove that the second player has a strategy which allows him to win the back-and-forth game of length n between these two models. By Fact 1, this will show that $\mathrm{QR}_{\mathcal{O}vs}(\mathrm{Even},N_n+1)>n$. As $N_p=\prod_{i=0}^{p}(2^i+1)\leqslant 2^{(p+1)(p+2)/2}$, this will give the desired lower bound $\mathrm{QR}_{\mathcal{O}vs}(\mathrm{Even},n)=\Omega(\sqrt{\log n})$.

Let us choose our first model $(V,\mathcal{M})$ and introduce some notations. Let $n_v=N_{n+1}$. Let V be $\mathbb{Q}^{n_v}$ ordered by the lexicographic order. For $1\leqslant i\leqslant n_v$, let $\varepsilon_i=(0,\ldots,0,1,0,\ldots,0)\in V$ where the element "1" is in position n_v-i+1. Let $\mathcal{M}=\{\varepsilon_1,\ldots,\varepsilon_{n_v}\}$. Observe that $\mathcal{M}$ is a basis of the $\mathbb{Q}$-vector space V. For $x\in V$, let $|x|$ denote $-x$ if $x<0$ and x otherwise. For $a,b\in V$, we say that $a\ll b$, or $a=o(b)$, if for all $n\in\mathbb{N}$, $n|a|\leqslant|b|$. Note that $0<\varepsilon_1\ll\ldots\ll\varepsilon_{n_v}$. Any element $v\in V$ can be written in a unique way $v=\sum_{i=1}^{r}\alpha_i a_i$ with $\alpha_i\in\mathbb{Q}\setminus\{0\}$, $a_i\in\mathcal{M}$ and $a_1\gg\ldots\gg a_r$. We use the following notations: $\mathrm{supp}(v)=\{a_1,\ldots,a_r\}$, $\mathrm{supp}(v,l)=\{a_i,\ i\leqslant\min(l,r)\}$, $z(v,j)=a_{\min(2^j,r)}$ and $\mathrm{T}_j(v)=\sum_{i=1}^{\min(2^j,r)}\alpha_i a_i$. Thus $z(v,j)=z(\mathrm{T}_j(v),j)$. Let us remark that if $|\mathrm{supp}(\mathrm{T}_j(x))|<2^j$ then $x=\mathrm{T}_j(x)$.

The second model $(W,\mathcal{N})$ is defined in the same way (we shall use the notations defined above in this model too). Let $n_w=N_{n+1}+1$; let W be $\mathbb{Q}^{n_w}$ with the lexicographic order. Let $\mathcal{N}$ be the canonical basis of W : thus $\mathcal{N}=\{\eta_1,\ldots,\eta_{n_w}\}$ and we have $0<\eta_1\ll\ldots\ll\eta_{n_w}$.

Let π denote the canonical projection from $\mathcal{M}\times\mathcal{N}$ onto $\mathcal{M}$. Given $R\subseteq\mathcal{M}\times\mathcal{N}$ a one-to-one function from a subset of $\mathcal{M}$ in $\mathcal{N}$, let $\mathcal{L}_R$ denote the linear mapping defined on $\mathrm{Vect}(\pi(R))$ and extending R. When it remains j steps to do, we shall have an isomorphism φ_{j+1} defined from E_{j+1} onto F_{j+1}. For the finite ordered sets $\mathcal{M}$ and $\mathcal{N}$ (the order being induced by the one of V or W), we shall make use of the distances d_∞ and d_j defined previously (thus, for $a < b$ two elements of $\mathcal{M}$, $d_\infty(a, b) = |\{x \in \mathcal{M}, a < x \leqslant b\}|$).

At the beginning of the game, we set $E_{n+1} = \mathrm{Vect}(\{\varepsilon_1, \varepsilon_{n_v}\})$, $F_{n+1} = \mathrm{Vect}(\{\eta_1, \eta_{n_w}\})$ and φ_{n+1} is the linear mapping from E_{n+1} onto F_{n+1} such that $\varphi(\varepsilon_1) = \eta_1$ and $\varphi(\varepsilon_{n_v}) = \eta_{n_w}$. We also set $R_{n+1} = \{(\varepsilon_1, \eta_1), (\varepsilon_{n_v}, \eta_{n_w})\} \subseteq \mathcal{M} \times \mathcal{N}$. At each step we shall maintain the following property $\mathcal{P}_j$.

a) For all $x, y \in \pi(R_j)$, $d_j(x, y) = d_j(R_j(x), R_j(y))$.
b) For all $v \in E_j$, we have $\mathrm{T}_j(\varphi_j(v)) = \mathcal{L}_{R_j}(\mathrm{T}_j(v))$. Similarly, for all $w \in F_j$, $\mathrm{T}_j(\varphi_j^{-1}(w)) = \mathcal{L}_{R_j^{-1}}(\mathrm{T}_j(w))$.
c) The mapping φ_j is an isomorphism of ordered $\mathbb{Q}$-vector space from E_j onto F_j such that, for all $x \in E_j$, $x \in \mathcal{M}$ iff $\varphi_j(x) \in \mathcal{N}$.

Let us remark that point b) means that $\mathcal{L}_{R_j}(\mathrm{T}_j(v))$ makes sense, so it implies $\mathrm{supp}(\mathrm{T}_j(v)) \subseteq \pi(R_j)$. Let us also remark that, as a consequence of a), R_j is a strictly increasing mapping from $\pi(R_j) \subseteq \mathcal{M}$ to $\mathcal{N}$. Let us show that $\mathcal{P}_{n+1}$ holds: a) comes from $|\mathcal{M}|, |\mathcal{N}| \geqslant N_{n+1}$, the other points are clear.

Let us assume that $n - j$ steps of the back-and-forth game have been carried out. It remains $j \geqslant 1$ steps to play. The isomorphism φ_{j+1} is defined from E_{j+1} onto F_{j+1}. By symmetry, we can assume that point v is chosen in $(V, \mathcal{M})$. Without loss of generality, we assume that $v \notin E_{j+1}$. Let $u \in \mathrm{Vect}(E_{j+1} \cup \{v\}) \setminus E_{j+1}$ *such that $z(u, j)$ is minimal* with respect to the order on V. Let $S = \mathrm{supp}(u, 2^j) \setminus \pi(R_{j+1})$. Thanks to Lemma 2, we now define the relation R_j extending R_{j+1} such that $\pi(R_j) = \pi(R_{j+1}) \cup S$.

Let $E_j = \mathrm{Vect}(E_{j+1} \cup \{u\})$. Let φ_j be the linear mapping (defined on E_j) extending φ_{j+1} and such that $\varphi_j(u) = \mathcal{L}_{R_j}(\mathrm{T}_j(u))$. Let $F_j = \varphi_j(E_j)$. In what follows, φ_j will be denoted φ, R_j will be denoted R and $\mathcal{L}_{R_j}$ sometimes denoted $\mathcal{L}_j$. Let us show we have $\mathcal{P}_j$. Let us first remark that φ is a linear mapping from E_j onto F_j. Let us show that φ is one to one. Let $w \in E_j$, $\varphi(w) = 0$. We write $w = \alpha u + e$ with $e \in E_{j+1}$ and $\alpha \in \mathbb{Q}$. If $\alpha = 0$, then $e = 0$ because φ_{j+1} is one to one. Let us suppose $\alpha \neq 0$. Thus $\varphi(e) = \varphi_{j+1}(e) = -\alpha\varphi(u) = -\alpha\mathcal{L}_j\mathrm{T}_j(u)$. Thanks to $\mathcal{P}_{j+1}$ *b)* for φ_{j+1}^{-1} we obtain $\mathrm{T}_{j+1}(e) = -\alpha\mathcal{L}_{R_{j+1}^{-1}}\mathrm{T}_{j+1}\mathcal{L}_j\mathrm{T}_j u$. Therefore $\mathrm{T}_{j+1}(e) = -\alpha\mathcal{L}_{R_{j+1}^{-1}}\mathcal{L}_j\mathrm{T}_j u = -\alpha\mathrm{T}_j(u)$ because this expression makes sense and R_j extends R_{j+1}. But $2^j < 2^{j+1}$, so $e = -\alpha\mathrm{T}_j(u)$. Now if $u \neq \mathrm{T}_j(u)$, this gives $w = e + \alpha u \notin E_{j+1}$ with $w = o(z(u, j))$ which is impossible by the choice of u. As $u = \mathrm{T}_j(u)$, we have $e = -\alpha u$ that is to say $w = 0$.

Point a) stems from construction. Let us show point b) for φ. Let $v \in E_j$. If $v \in E_{j+1}$, it is clear by $\mathcal{P}_{j+1}$ since $2^{j+1} \geqslant 2^j$, $R_{j+1} \subseteq R_j$ and φ extends φ_{j+1}. Hence we suppose $v \notin E_{j+1}$ Thus $v = \alpha u + e$ where u is the vector chosen above, $\alpha \in \mathbb{Q} \setminus \{0\}$ and $e \in E_{j+1}$. The following holds.

$$\mathrm{T}_j(v) = \mathrm{T}_j(\alpha \mathrm{T}_j(u) + \mathrm{T}_{j+1}(e)) \ . \tag{1}$$

Proof:
i) Let us suppose $z(\mathrm{T}_{j+1}(e)) \leqslant z(\mathrm{T}_j(u))$. Thus $\mathrm{T}_{j+1}(e) = e + o(z(\mathrm{T}_j(u)))$ and $v = \alpha u + e = \alpha \mathrm{T}_j(u) + \mathrm{T}_{j+1}(e) + o(z(\mathrm{T}_j(u)))$. As $z(\mathrm{T}_j(v)) \geqslant z(\mathrm{T}_j(u))$ by the choice of u, we obtain relation (1) by truncating the previous equality at the order 2^j.
ii) Now let us suppose $z(\mathrm{T}_{j+1}(e)) \geqslant z(\mathrm{T}_j(u))$.
ii-a) If $\mathrm{T}_{j+1}(e) = e$, in particular we have $\mathrm{T}_{j+1}(e) = e + o(z(\mathrm{T}_j(u)))$ and we finish as previously.
ii-b) Otherwise, $|\mathrm{supp}(e, 2^{j+1})| = 2^{j+1}$. Moreover $v = e + \alpha u = \mathrm{T}_{j+1}(e) + \alpha \mathrm{T}_j(u) + o(z(\mathrm{T}_{j+1}(e)))$. As the sum $\mathrm{T}_{j+1}(e) + \alpha \mathrm{T}_j(u)$ has at least 2^j terms from $\mathrm{T}_{j+1}(e)$, we obtain (1) by truncating the previous equality at the order 2^j.

The following also holds.

$$\mathrm{T}_j(\varphi(v)) = \mathrm{T}_j(\alpha \mathrm{T}_j(\varphi(u)) + \mathrm{T}_{j+1}(\varphi(e))) \ . \tag{2}$$

Proof:
i) Let us suppose $|\mathrm{supp}(e, 2^{j+1})| < 2^{j+1}$. Thus $e = \mathrm{T}_{j+1}(e)$. By $\mathcal{P}_{j+1}$, we obtain $\mathrm{T}_{j+1}(\varphi(e)) = \mathcal{L}_{j+1}(\mathrm{T}_{j+1}(e)) = \mathcal{L}_{j+1}(e)$. But $\mathcal{L}_{j+1}(e)$ has strictly less than 2^{j+1} terms so $\mathrm{T}_{j+1}(\varphi(e)) = \varphi(e)$. Let us recall that $\varphi(u) = \mathrm{T}_j(\varphi(u))$ by the choice of $\varphi(u)$. By substituting these terms in $\mathrm{T}_j(\varphi(v)) = \mathrm{T}_j(\alpha\varphi(u) + \varphi(e))$ we obtain (2).
ii) Otherwise $|\mathrm{supp}(e, 2^{j+1})| = 2^{j+1}$. Thus $\mathrm{T}_{j+1}(\varphi(e)) = \mathcal{L}_{j+1}(\mathrm{T}_{j+1}(e))$ has 2^{j+1} terms. But $\varphi(v) = \alpha\varphi(u) + \varphi(e) = \alpha \mathrm{T}_j(\varphi(u)) + \mathrm{T}_{j+1}(\varphi(e)) + o(z(\mathrm{T}_{j+1}(\varphi(e))))$. As $\alpha \mathrm{T}_j(\varphi(u)) + \mathrm{T}_{j+1}(\varphi(e))$ has at least 2^j terms from $\mathrm{T}_{j+1}(\varphi(e))$, we obtain (2) by truncating the previous equality at the order 2^j.

Now let us prove P_j *b)* for φ. Let $v \in E_j$. We write $v = \alpha u + e$ with $e \in E_{j+1}$ and $\alpha \in \mathbb{Q}$. By (2), $\mathrm{T}_j(\varphi(v)) = \mathrm{T}_j(\alpha \mathrm{T}_j(\varphi(u)) + \mathrm{T}_{j+1}(\varphi(e)))$. But $\mathrm{T}_j(\varphi(u)) = \varphi(u) = \mathcal{L}_j \mathrm{T}_j(u)$ by the choice of $\varphi(u)$. Moreover, by $\mathcal{P}_{j+1}$, $\mathrm{T}_{j+1}(\varphi(e)) = \mathcal{L}_{j+1}\mathrm{T}_{j+1}(e)$. And $\mathcal{L}_{j+1}\mathrm{T}_{j+1}(e) = \mathcal{L}_j \mathrm{T}_{j+1}(e)$ since $\mathcal{L}_j$ extends $\mathcal{L}_{j+1}$. By the linearity of $\mathcal{L}_j$, this gives $\mathrm{T}_j(\varphi(v)) = \mathrm{T}_j(\mathcal{L}_j(\alpha \mathrm{T}_j(u) + \mathrm{T}_{j+1}(e)))$. Clearly, if $\mathrm{T}_j \mathcal{L}_j(x)$ makes sense for $x \in E_j$, then $\mathcal{L}_j \mathrm{T}_j(x) = \mathrm{T}_j \mathcal{L}_j(x)$. Thus we have $\mathrm{T}_j(\varphi(v)) = \mathcal{L}_j \mathrm{T}_j(\alpha \mathrm{T}_j(u) + \mathrm{T}_{j+1}(e))$. With relation (1) we obtain $\mathrm{T}_j(\varphi(v)) = \mathcal{L}_j(\mathrm{T}_j(v))$.

We now show point b) for φ^{-1}. Let $w \in F_j$ and $v \in E_j$ such that $w = \varphi(v)$. We have $\mathrm{T}_j(w) = \mathrm{T}_j(\varphi(v)) = \mathcal{L}_j(\mathrm{T}_j(v))$ by $\mathcal{P}_j$ *b)* for φ. Moreover, $\mathcal{L}_{R_j}^{-1} = \mathcal{L}_{R_j^{-1}}$; therefore $\mathrm{T}_j(\varphi^{-1}(w)) = \mathcal{L}_{R_j^{-1}}(\mathrm{T}_j(w))$. This proves point b) of $\mathcal{P}_j$ for φ^{-1}.

It remains to prove c). If $a \in E_j \cap \mathcal{M}$, then by $\mathcal{P}_j$ b) we have $\mathrm{T}_j(\varphi(a)) = \mathcal{L}_j \mathrm{T}_j(a) = \mathcal{L}_j(a) = R(a)$. But $|\mathrm{supp}(R(a))| = 1 < 2^j$, so $\varphi(a) = R(a) \in \mathcal{N}$. In the same way, if $x \in E_j$ is positive, then $x = \alpha a + o(a)$ with $a \in \mathcal{M}$ and $\alpha > 0$. By point b) of $\mathcal{P}_j$, we have $\varphi(a) = \alpha R_j(a) + o(R_j(a))$. But $R_j(a) \in \mathcal{N}$; thus $R_j(a) > 0$, which proves $\varphi(x) > 0$. The same works for φ^{-1}, so it completes the proof of point c). This ends the back-and-forth game.□

Does a similar result hold in real-closed fields? We have a weaker bound in o-minimal structures having quantifier elimination: see Corollary 5. We now show an upper bound.

The rough idea is this one. To express that $|\mathcal{I} \cap]a, b[| \geqslant 2^{p^2}$, it is enough to have a set S of 2^{2p} elements of $\mathcal{I}$ such that between two consecutive elements of S, there are at least $2^{(p-1)^2}$ elements of $\mathcal{I}$ (this will be done by induction). The set S is represented by the sum of its elements, from which it is possible to extract the elements with a formula of quantifier depth p. However, if several sets of elements give the same sum, there are no canonical elements to extract from the sum, and intervals considered in the recursion step can overlap. That is why we make sure no other set gives the same sum by weighting the coefficients of the elements in the sum.

Proposition 3 $\mathrm{QR}_{Ovs}(\mathrm{Card}_m, n) = O(\sqrt{\log m})$.

Proof. We shall work over $\mathbb{Q}$. Let n be fixed. In the following we assume that $|\mathcal{I}| \leqslant n$. We define the sequence m_i by $m_0 = 1$ and $m_p = 2^p + (2^p + 1)m_{p-1}$. We also need to define some families of formulas. For $k \geqslant 1$ and $\bar{\alpha} = (\alpha_1, \ldots, \alpha_k) \in \mathbb{N}^k$, we build a formula $S_{\bar{\alpha}}(a, b, x)$ of quantifier rank $O(\log k)$ which is true if and only if

$$\exists x_1, \ldots, x_k \in \mathcal{I} \; a < x_1 < x_2 < \ldots < x_k < b \wedge x = \sum_{i=1}^{k} \alpha_i x_i \; .$$

We also build $E_{(\alpha_1, \ldots, \alpha_k), j}(a, b, x, z)$ for $1 \leqslant i \leqslant k$, with quantifier rank $O(\log k)$ too, which is true if and only if

$$\exists x_1, \ldots, x_k \in \mathcal{I} \; a < x_1 < x_2 < \ldots < x_k < b \wedge x = \sum_{i=1}^{k} \alpha_i x_i \wedge x_j = z \; .$$

The construction of these formulas is quite obvious and not detailed here.

We shall prove by induction on p that for any $m \leqslant m_p$ there exists a formula $F_m(a,b)$ of quantifier rank $O(p)$ expressing that $|\mathcal{I} \cap]a,b[| \geqslant m$. This is clear for $p = 0$. We now want to show the property for rank p assuming it is true for $p-1$. Let $m_{p-1} < m \leqslant m_p$; then $m = (m_{p-1}+1)q + r$ for some $1 \leqslant q \leqslant 2^p$ and $r \leqslant m_{p-1}$. Let $\bar{\alpha} = (\alpha_1, \ldots, \alpha_q) \in \mathbb{N}^q$. We define

$$\begin{aligned} G^m_{\bar{\alpha}} := &\exists x S_{\bar{\alpha}}(a,b,x) \wedge \exists z E_{\bar{\alpha},1}(a,b,x,z) \wedge F_{m_{p-1}}(a,z) \wedge \\ &\bigwedge_{i=1}^{q-1} \exists z_1, z_2 E_{\bar{\alpha},i}(a,b,x,z_1) \wedge E_{\bar{\alpha},i+1}(a,b,x,z_2) \wedge F_{m_{p-1}}(z_1,z_2) \wedge \\ &\exists z E_{\bar{\alpha},q}(a,b,x,z) \wedge F_r(z,b). \end{aligned}$$

Let $N = n^{2q}$ and $\mathcal{A}$ a set of $Nq+1$ elements of $\mathbb{N}^q$ in general position. We claim that

$$(\mathbb{Q}, \mathcal{I}) \models \bigwedge_{\bar{\alpha} \in \mathcal{A}} G^m_{\bar{\alpha}}(a,b) \iff |\mathcal{I} \cap]a,b[| \geqslant m.$$

Proof of the claim. From right to left: let us assume $|\mathcal{I} \cap]a,b[| \geqslant m$. Let $x_1 < x_2 < \ldots < x_m$ be some elements of $\mathcal{I} \cap]a,b[$. Let $\bar{\alpha}$ be any element in $\mathcal{A}$. Then $G^m_{\bar{\alpha}}$ is true – take $x = \sum_{i=1}^q \alpha_{i(m_{p-1}+1)} x_{i(m_{p-1}+1)}$ for the first quantifier, and $z = x_i$ when asked for an element z such that $E_{\bar{\alpha},i}(a,b,x,z)$. Thus $\bigwedge_{\bar{\alpha} \in \mathcal{A}} G^m_{\bar{\alpha}}$ is true.

Conversely, let us assume that $\bigwedge_{\bar{\alpha} \in \mathcal{A}} G^m_{\bar{\alpha}}$ is true. Let $x_1 < x_2 < \ldots < x_l$ denote the elements of $\mathcal{I} \cap]a,b[$, where $0 \leqslant l \leqslant n$ by hypothesis. Let us define a family of hyperplanes $\mathcal{H}_{\bar{x}}$ which depends on $(x_1, \ldots, x_l)$. The family $\mathcal{H}_{\bar{x}}$ is composed of all hyperplanes of the form

$$\sum_{i=1}^q x_{f(i)} A_i = \sum_{i=1}^q x_{g(i)} A_i$$

where f and g runs over all pairs of strictly increasing functions from $\{1, \ldots, q\}$ to $\{1, \ldots, l\}$ with $f \neq g$. As $|\mathcal{H}_{\bar{x}}| \leqslant (l^q)^2 \leqslant N$, we have $|\mathcal{A}| > q|\mathcal{H}_{\bar{x}}|$, so there must be $\bar{\alpha} \in \mathcal{A}$ which does not lie on any hyperplane of $\mathcal{H}_{\bar{x}}$. The fact that $G^m_{\bar{\alpha}}(a,b)$ is true implies that $|\mathcal{I} \cap]a,b[| \geqslant m$. There are indeed some elements $z_1 < \ldots < z_q$ in $\mathcal{I} \cap]a,b[$ such that $x = \sum_{i=1}^q \alpha_i z_i$ works for the first quantifier in $G^m_{\bar{\alpha}}(a,b)$. Moreover, any other such sequence $z'_1 < \ldots < z'_q$ is such that $\sum_{i=1}^q \alpha_i z'_i \neq x$, because $\bar{\alpha}$ is not on $\mathcal{H}_{\bar{x}}$. It means that, for any $1 \leqslant i \leqslant q$, the only z such that $E_{\bar{\alpha},i}(a,b,x,z)$ is z_i. By induction hypothesis, we can deduce that $|\mathcal{I} \cap]z_i, z_{i+1}[| \geqslant m_{p-1}$ for $1 \leqslant q-1$, $|\mathcal{I} \cap]a, z_1[| \geqslant m_{p-1}$ and at last $|\mathcal{I} \cap]z_q, b[| \geqslant r$. Thus $|\mathcal{I} \cap]a,b[| \geqslant m$. □

Corollary 2 $\mathrm{QR}_{\mathcal{O}vs}(\mathrm{Even}, n) = \Theta(\sqrt{\log n})$. □

Proposition 4 *In an ordered $\mathbb{Q}$-vector space, we can express that $|\mathcal{I}| \geqslant m$ with a formula of quantifier rank $O(\sqrt{\log m})$.*

Proof. Compared to the previous proposition, we no longer have a bound on $|\mathcal{I}|$. However we claim that a formula, built as in the previous proposition, expressing $|\mathcal{I}| \geqslant m$ whenever $|\mathcal{I}| \leqslant m-1$ works even if we remove the assumption that $|\mathcal{I}| \leqslant m-1$. Indeed, the only way this formula can give the wrong answer is in saying that $|\mathcal{I}| \geqslant m$. But this can only happen when $|\mathcal{I}| \geqslant m$ because the formula works when $|\mathcal{I}| \leqslant m-1$. So this formula never goes wrong. □

Remark 2 *Is it possible to express that a finite set $\mathcal{I}$ embedded in a $\mathbb{Q}$-vector space has cardinal at least m with a formula of quantifier rank $O(\log m)$? – compared to Theorem 2, we do not have any bound on $\mathcal{I}$ anymore.*

5 Connectivity of embedded graphs

In this section we consider a finite graph G embedded in an infinite structure M. Thus we shall add two predicates to the signature of M: a unary predicate V which interprets the vertices $\mathcal{V}$ of the embedded graph, and a binary one E for the edges. We shall use $d(\cdot,\cdot)$ for the distance in the graph.

We are interested in connectivity and reachability. The query Connected asks if the graph G is connected. The query Reach_m will have two free variables a and b and will be true if $a, b \in \mathcal{V}$ and $d(a,b) \leqslant m$. The query Reach is defined in the same way, except there is no bound on the length of the path anymore. Once again we shall consider restriction of these queries to the case where we have a bound on $|\mathcal{V}|$. Thus $\text{QR}_M(\text{Connected}, n)$ is the smallest quantifier rank possible for a formula expressing that a graph embedded in M is connected, assuming that is has at most n vertices. Of course the result of Lemma 1 holds for these queries too: if two structures M and M' are elementarily equivalent, then $\text{QR}_M(\text{Connected}, n) = \text{QR}_{M'}(\text{Connected}, n)$ and $\text{QR}_M(\text{Reach}_m, n) = \text{QR}_{M'}(\text{Reach}_m, n)$. Let us notice that $\text{QR}_M(\text{Reach}, n) = \text{QR}_M(\text{Reach}_{n-1}, n)$. Another remark is that $\forall a, b\ V(a) \wedge V(b) \rightarrow \text{Reach}(a,b)$ expresses the connectivity, so $\text{QR}_M(\text{Connected}, n) \leqslant \text{QR}_M(\text{Reach}, n) + 2$. We can obtain a result similar to Theorem 3 for a finite graph embedded in an ordered $\mathbb{Q}$-vector space.

Corollary 3 $\text{QR}_{\mathcal{O}vs}(\text{Connected}, n) = \Omega(\sqrt{\log n})$.

Proof. We use a usual first-order reduction from parity to connectivity. Let $\mathcal{I}$ be a set composed of the elements $v_1 < v_2 < \ldots < v_n$. We consider the graph $G_n = (V, E)$ over $V = \{v_1, \ldots, v_n\}$ where $E(v_i, v_j)$ holds iff $|i-j| = 2$ or $\{i,j\} = \{1,n\}$. For $n \geqslant 2$, G_n is connected iff $n = |\mathcal{I}|$ is even. As we can express E with a formula of quantifier rank 2 (with the help of predicate I interpreting $\mathcal{I}$) in any ordered structure M, we obtain $\text{QR}_M(\text{Even}, n) \leqslant$

$\mathrm{QR}_M(\mathrm{Connected}, n) + 2$. It remains to apply this to the theory $\mathcal{O}vs$. □

Using techniques similar to the previous ones, we can establish a lower bound for algebraically closed fields. By the remark made at the beginning of this section, it is sufficient to prove this result in $\mathbb{C}$. Recall from section 4 the definition of N_i for $i \geqslant 0$. Let us call C_n the cycle of length N_{n+1}. Let G_n be the graph C_n and H_n be the graph composed of two disjoint copies of C_n. As in section 4 we define d_j to be the truncature of d (the distance in the graph) relative to N_j. We begin with an analogue of Lemma 2.

Lemma 3 *Consider the variation of the back-and-forth game between two finite graphs (over a domain without order) where it is possible to choose up to 2^{j-1} elements at once, on the same side, when it remains exactly j moves to play. Then the second player has a strategy to win this modified game of length n between G_n and H_n.*

Proof. Let V_G (resp. V_H) be the vertices of G_n (resp. H_n). When it remains j moves to play, there is a partial isomorphism φ_j from $A_j \subset V_G$ onto $B_j \subset V_H$. In order to show the Lemma, it is enough to maintain the following property

$$\forall x, y \in A_j, d_j(x, y) = d_j(\varphi_j(x), \varphi_j(y)) \ .$$

This is done in a way very similar to Lemma 2. □

Proposition 5 $\mathrm{QR}_{ACF_0}(\mathrm{Connected}, n) = \Omega(\sqrt{\log n})$.

Proof. By the remark made at the beginning of this section, it is sufficient to prove this result in $\mathbb{C}$. Let us consider $\mathcal{G} = (V_\mathcal{G}, E_\mathcal{G})$ an embedding of G_n in $\mathbb{C}$ such that all vertices $V_\mathcal{G}$ are algebraically independent over $\mathbb{Q}$. In the same way, let $\mathcal{H} = (V_\mathcal{H}, E_\mathcal{H})$ be an embedding of H_n in $\mathbb{C}$ such that all vertices $V_\mathcal{H}$ are algebraically independent over $\mathbb{Q}$. We only have to prove that the second player can win the game of length n between $(\mathbb{C}, \mathcal{G})$ and $(\mathbb{C}, \mathcal{H})$. Then the conclusion comes from fact 1.

The back-and-forth game is done in a way similar to Theorem 1. When it remains j steps to do, E_j denotes the space where φ is defined and $F_j = \varphi(E_j)$. Let $E_n = F_n = \overline{\mathbb{Q}}$ and $\varphi = \mathrm{Id}_{\overline{\mathbb{Q}}}$. At each step we shall maintain the following property $\mathcal{P}_j$.

a) The mapping φ_j is an isomorphism of 0-characteristic algebraically closed fields from E_j onto F_j.
b) If there exists $a \in V_{\mathcal{G}} \setminus E_j$ such that $a \in \overline{E_j \cup A}$ and $A \subset V_{\mathcal{G}} \setminus (E_j \cup \{a\})$, then $|A| \geqslant 2^{j-1}$. And the corresponding property in $(\mathbb{C}, \mathcal{H})$.
c) $\forall x \in E_j, x \in V_{\mathcal{G}}$ iff $\varphi_j(x) \in V_{\mathcal{H}}$.
d) $\forall x, y \in E_j, d_j(x, y) = d_j(\varphi_j(x), \varphi_j(y))$.

Observe that a consequence of d) is that $\forall x, y \in E_j,\ (x, y) \in E_{\mathcal{G}}$ if and only if $(\varphi_j(x), \varphi_j(y)) \in E_{\mathcal{H}}$. Let us suppose that $n - j - 1$ steps have been done. The isomorphism φ is defined on E_{j+1} and it remains $j + 1$ steps to do. Property $\mathcal{P}_{j+1}$ is satisfied by induction hypothesis. By symmetry, we can assume that the point is chosen in $(\mathbb{C}, \mathcal{G})$. This point will be denoted v. We can also assume $v \notin E_{j+1}$. There are two cases:

• First case: $v \in \overline{E_{j+1} \cup \{a_1, \ldots, a_r\}}$ with $a_i \in V_{\mathcal{G}} \setminus E_{j+1}$ distinct and $r \leqslant 2^{j-1}$. Then the second player will choose some elements $b_1, \ldots, b_r$ in $V_{\mathcal{H}} \setminus F_{j+1}$ *as suggested in Lemma 3*, and we define $\varphi(a_i) = b_i$. What does it mean exactly? From the move v played by the first player in the game between $(\mathbb{C}, \mathcal{G})$ and $(\mathbb{C}, \mathcal{H})$, the second player finds some $\{a_1, \ldots, a_r\}$ that correspond to a move in the modified version of back-and-forth game defined in Lemma 3. The second player chooses some points $\{b_1, \ldots, b_r\}$ according to the strategy described in Lemma 3, and plays according to this move in the game between $(\mathbb{C}, \mathcal{G})$ and $(\mathbb{C}, \mathcal{H})$.
• Second case: let $f \notin \overline{F_{j+1} \cup V_{\mathcal{H}}}$. Let $\varphi(v) = f$. We set $E_j = \overline{E_{j+1} \cup \{v\}}$.

Property a), b) and c) hold since the game is played as in Theorem 1. Property d) hold because we chose points as described in Lemma 3. □

We now establish an upper bound for reachability in $\mathbb{Q}$-vector spaces – which provides the same bound for theories $\mathcal{O}vs$ and ACF_0.

Proposition 6 $\mathrm{QR}_{\mathbb{Q}vs}(\mathrm{Reach}_m, n) = O(\sqrt{\log m})$.

Proof. Let n be fixed. We define the sequence m_i by $m_0 = 1$ and $m_{p+1} = (2^p + 1)m_p$. Let us show by induction on p that, for any $m_p < m \leqslant m_{p+1}$, there is a formula of $\mathbb{Q}$-vector space with quantifier rank $O(p)$ expressing "$d(\cdot, \cdot) \leqslant m$" if $|\mathcal{V}| \leqslant n$. Let $m = m_p q + r$, with $1 \leqslant q \leqslant 2^p + 1$ and $0 \leqslant r < m_p$. Now, for two vertices u and v, $d(u, v) \leqslant m$ if and only if there exists some vertices $u = c_0, c_1, c_2, \ldots, c_q$ such that $d(c_i, c_{i+1}) \leqslant m_p$, and $d(c_q, v) \leqslant r$. The sequence $(c_1, c_2, \ldots, c_q)$ will be represented by a weighted sum, from which it is possible to retrieve the c_i with formulas of quantifier rank $O(p)$: this is done as in the previous result. The statements concerning $d(c_i, c_{i+1})$ are handled in a recursive way, in parallel for all i. □

Corollary 4 *For* $T \in \{\mathbb{Q}vs, ACF_0, \mathcal{O}vs\}$, $\left| \begin{array}{l} \mathrm{QR}_T(\mathrm{Reach}_m, n) = \Theta(\sqrt{\log m}) \\ \mathrm{QR}_T(\mathrm{Reach}, n) = \Theta(\sqrt{\log n}) \\ \mathrm{QR}_T(\mathrm{Connected}, n) = \Theta(\sqrt{\log n}) \end{array} \right.$

□

6 Relationship with active-natural collapse

Consider a L-structure M expanded with a unary predicate I. We shall restrict ourselves to the case where I interprets a finite set $\mathcal{I}$. We recall that in an active formula, quantifiers are of the type $\exists x \in \mathcal{I}$ and $\forall x \in \mathcal{I}$; quantifiers ranging over the whole universe are not allowed.

Definition 1 *A L-structure M is said to have the active-natural collapse (for one unary predicate) if, for any first-order formula ϕ over $L \cup \{I\}$, there exists a first-order active formula ψ over $L \cup \{I\}$ such that for any finite set $\mathcal{I}$, we have $(M, \mathcal{I}) \models \phi \leftrightarrow \psi$.*

On a structure M with active-natural collapse, there is a relationship between how the quantifier rank grows when transforming a natural semantics formula into an equivalent one in active semantics, and the quantifier rank needed to express parity. This is detailed in the next proposition.

Let M be a structure having the active-natural collapse property for one unary predicate. For any first-order formula ψ over $L \cup \{I\}$, let us define $a_M(\psi)$ to be be the smallest quantifier rank for an active formula equivalent to ψ. Let $\alpha_M(n) = \sup\{a_M(\psi), \mathrm{qr}(\psi) = n\}$.

Proposition 7 *For any structure M having the active-natural collapse property, $\alpha_M(\mathrm{QR}_M(\mathrm{Even}, n) + 1) \geqslant \lfloor \log n \rfloor + 1$. Moreover, if M is stable, we have the stronger inequality $\alpha_M(\mathrm{QR}_M(\mathrm{Even}, n) + 1) \geqslant n$.*

Proof. Let ψ be a formula expressing parity of $|\mathcal{I}|$ for $|\mathcal{I}| \leqslant n$. Let us consider a structure M', elementarily equivalent to M, that contains a sequence of indiscernibles $E = \{e_i,\ i \in \mathbb{N}\}$ (such a model exists by the Ehrenfeucht-Mostowski theorem (4)). The same formula ψ still works in M'. Let ψ' be the formula obtained by replacing all sub-formulas $I(t)$ by $\forall z\ ((z = t) \rightarrow I(z))$ in ψ, where z is a new variable. Note that $\mathrm{qr}(\psi') \leqslant \mathrm{qr}(\psi) + 1$. Let ψ_{act} be an active semantics formula equivalent to ψ' and such that $\mathrm{qr}(\psi_{act}) \leqslant \alpha_M(\mathrm{qr}(\psi'))$. Now when we restrict ourselves to the case where $\mathcal{I} \subset E$, the formula ψ_{act} is equivalent to a pure order formula ψ_o with $\mathrm{qr}(\psi_o) = \mathrm{qr}(\psi_{act})$. Applying the bound in the pure ordered case recalled in the introduction, we obtain $\mathrm{qr}(\psi_o) \geqslant \lfloor \log n \rfloor + 1$. This leads to the desired bound.

Now if M is stable, then E is a set of indiscernibles and we can apply the pure case bound for parity. □

Corollary 5 *Let M be an o-minimal structure that admits quantifier elimination. Then* $\mathrm{QR}_M(\mathrm{Even}, n) \geqslant \log\log n - \Theta(1)$.

Proof. The active-natural collapse algorithm described in (3) allows to obtain an active semantics formula ψ_{act} equivalent to a given formula ψ such that $\mathrm{qr}(\psi_{act}) \leqslant 2^{qr(\psi)} + O(1)$ (as mentioned there, there is no need for ψ to be in prenex form to apply this algorithm). Thus $\alpha_M(n) \leqslant 2^n + O(1)$. It just remains to apply the previous proposition.□

Question 1 *We have proved* $\Omega(\log\log n) \leqslant \mathrm{QR}_{\mathbb{R}}(\mathrm{Even}, n) \leqslant O(\sqrt{\log n})$, *where* $\mathbb{R}$ *stands for* $(\mathbb{R}, +, -, \times, <)$. *Is it possible to make some back-and-forth games in a real closed field to improve this lower bound?*

As parity up to n is expressible with logarithmic quantifier rank over a $\mathbb{Q}$-vector space by Theorem 2, Proposition 7 shows that one cannot avoid an exponential growth of the quantifier rank when transforming natural semantics formulas of $\mathbb{Q}$-vector spaces into equivalent ones in active semantics: $\alpha_{\mathbb{Q}vs}(n) \geqslant \Omega(2^n)$.

Acknowledgments

I would like to thank Luc Ségoufin and Pascal Koiran for useful comments and careful reading of the paper.

References

[1] J. T. Baldwin and M. Benedikt. Stability theory, permutations of indiscernibles, and embedded finite models. *Trans. Amer. Math. Soc.*, 352(11):4937–4969, 2000.

[2] M. Benedikt, G. Dong, L. Libkin, and L. Wong. Relational expressive power of constraint query languages. *Journal of the ACM*, pages 1–34, 1998.

[3] M. Benedikt and L. Libkin. Relational queries over interpreted structures. *Journal of the ACM*, 47(4):644–680, 2000.

[4] C. C. Chang and H. J. Keisler. *Model Theory.* North Holland, third edition, 1990.

[5] O. Chapuis and P. Koiran. Definability of geometric properties in algebraically closed fields. *Mathematical Logic Quarterly*, 45(4):533–550, 1999.

[6] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer-Verlag, 1995.

[7] D. Flath and S. Wagon. How to pick out the integers in the rationals: An application of number theory to logic. *American Mathematical Monthly*, 98:812–823, 1991.

[8] S. Grumbach and J. Su. Queries with arithmetical constraints. *Theoretical Computer Science*, 173:151–181, 1997.

[9] W. Hodges. *A Shorter Model Theory*. Cambridge University Press, 1997.

[10] N. Immerman. *Descriptive Complexity*. Graduate Texts in Computer Science. Springer, 1998.

[11] G. Kuper, L. Libkin, and J. Paredaens, editors. *Constraint Databases*. Springer-Verlag, 2000.

[12] J. Robinson. Definability and decision problems in arithmetic. *Journal of Symbolic Logic*, 14:98–114, 1949.