Sparse NP-complete problems over the reals with addition

Hervé Fournier

March 2000

Abstract

An analog of Mahaney's Theorem was shown, stating that there is no sparse NP-complete problem over the reals with addition and equality – here sparse is defined in terms of dimension. We extend this to the case of the Turing reduction, then turn our attention toward the ordered case where it is shown that such sparse Turing-complete languages do exist. At last, we formulate a conjecture concerning the case of the many-one reduction.

Keywords: sparse sets, real complexity, knapsack.

1 Introduction

The study of sparse NP-complete problems began at the end of the seventies. We will just recall some basic facts, a survey about recent results in this area can be found in [3]. A language $L \subset \{0,1\}^*$ is said to be sparse if $|L \cap \{0,1\}^n| = n^{O(1)}$. Among the first results, let us cite the Karp-Lipton theorem : if there is a sparse NP-hard language for the Turing reduction, then the polynomial hierarchy collapses to its second level. At the same time, Mahaney showed that if there is a sparse NP-hard language for the many-one reduction, then P = NP.

Cucker, Koiran and Matamala studied a similar question over the reals with addition and equality [4]. The real classes we will consider are to be understood in the Blum,Shub and Smale setting [2, 1, 6]. The structure $\mathbb{R}_{vs}$ is $(\mathbb{R}, =, 0, 1, +, -)$ while $\mathbb{R}_{ovs}$ is $(\mathbb{R}, \leq, 0, 1, +, -)$. Let us recall the definition of a sparse real language.

Definition 1 *A language $L \subset \mathbb{R}^\infty$ definable over $\mathbb{R}_{vs}$ or $\mathbb{R}_{ovs}$ is sparse if* $\dim(L \cap \mathbb{R}^n) = (\log n)^{O(1)}$.

It was shown in [4] that no $\mathbb{R}_{vs}$-definable sparse set can be $\mathrm{NP}_{\mathbb{R}_{vs}}$-complete for many-one reduction. Here we investigate similar questions. In section 2, we obtain a similar result in the case of the Turing reduction. In section 3, we deal with the reals with addition and order. There we exhibit a sparse

Turing-complete problem. The case of the many-one reduction over $\mathbb{R}_{ovs}$ remains unsettled.

2 Reals with equality

We prove here a result similar to the one obtained in [4] for the Turing reduction. Let us recall the definition of the real knapsack problem $\mathrm{KP}_{\mathbb{R}}$.

$$\mathrm{KP}_{\mathbb{R}} \cap \mathbb{R}^{n+1} = \{(y, x_1, \ldots, x_n),\ \exists I \subset \{1, \ldots, n\} \sum_I x_i = y\}$$

It is straightforward that $\mathrm{KP}_{\mathbb{R}}$ belongs to both $\mathrm{NP}_{\mathbb{R}_{ovs}}$ and $\mathrm{NP}_{\mathbb{R}_{vs}}$. The geometry of $\mathrm{KP}_{\mathbb{R}}$ is simple : $\mathrm{KP}_{\mathbb{R}} \cap \mathbb{R}^{n+1}$ is the union of 2^n hyperplanes. First we need to bound the number these hyperplanes containing a given linear subspace.

Lemma 1 *Let A be a linear subspace of dimension k in $\mathbb{R}^{n+1}$. Then there are at most 2^{n-k} hyperplanes of $\mathrm{KP}_{\mathbb{R}} \cap \mathbb{R}^{n+1}$ whose direction contains A.*

Let $\{v_1, \ldots, v_k\}$ be a base of A. We note V the matrix whose lines are $\{v_1, \ldots, v_k\}$. To a (column) vector $U = (u_1, \ldots, u_n) \in \{0,1\}^n$ corresponds the hyperplane with equation $\sum_i u_i x_i - y = 0$. This hyperplane contains A if $VU = 0$. Choosing k independant columns in V, this gives an invertible matrix V' and a system $V'U' = -V''U''$. Then one solution at most is found when $U'' \in \{0,1\}^{n-k}$ is fixed. □

Proposition 1 *No sparse $\mathbb{R}_{vs}$-definable language is $\mathrm{NP}_{\mathbb{R}_{vs}}$-hard for the Turing reduction.*

Let us suppose $\mathrm{NP}_{\mathbb{R}_{vs}} \leq_T S$ with S sparse and $\mathbb{R}_{vs}$-definable. Let φ be a $\mathrm{P}_{\mathbb{R}_{vs}}(S)$ algorithm deciding $\mathrm{KP}_{\mathbb{R}}$, and t a polynomial bounding its running time. We define the generic path γ for $\mathbb{R}^n$: each time the program reaches a branch (a test or a question to the oracle), it takes the only way followed by a set of input of dimension n. For a path α, we note X_α the set of inputs from $\mathbb{R}^n$ following this path and C_α its complementary. For a set $A \subseteq \mathbb{R}^n$, let $H(A)$ be the set of hyperplanes of KP_n included in $\overline{A}$ possibly after translation and $N(A) = |H(A)|$. We set $N_\alpha = |H(C_\alpha)|$. At last, we note D_α the set of points that do not follow the generic path at the test just after α.

Let α be a prefix of the generic path for $\mathbb{R}^n$. We note α' the generic path corresponding to an additional step of computation. If this step is an operation or a test, then D_α is empty or is a hyperplane, so $N(D_\alpha) \leq 1$. Let us consider a node made of a question to the oracle : the question is given by an affine function of the coordinates $x_1, \ldots, x_n$ of the input. This function only depends on α. Let L be the linear part of this function. Let

$A = L^{-1}(S)$. First case : $\dim A \leq n-1$. The generic path takes the branch labeled *no* and $D_\alpha = A+p$ for a point $p \in \mathbb{R}^n$. We will now bound $N(A)$. The function L is in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $m \leq t(n)$. We recall that $\dim S_n \leq a(\log n)^b$. If $\dim\ ImL > a(\log t(n))^b + 1$ then $\dim A \leq n-2$ and $N(A) = 0$. Otherwise $\dim\ ImL \leq a(\log t(n))^b + 1$, thus $\dim\ KerL \geq n - a(\log t(n))^b - 1$. All the hyperplanes of $H(A)$ contain $KerL$. By Lemma 1, the number of such hyperplanes is bounded by $2^{n-\dim(KerL)}$. Thus $N(A) \leq 2^{(\log n)^{O(1)}}$. Second case : $\dim A = n$, so $D_\alpha = A^c$. Since S sparse, we have $\dim\ ImL \leq a(\log t(n))^b$ and for the same reason as above $N(A^c) \leq 2^{(\log n)^{O(1)}}$. It remains to notice that for A and B both definable, $N(A \cup B) \leq N(A) + N(B)$. Thus $N_{\alpha'} = N(C_{\alpha'}) = N(C_\alpha \cup D_\alpha) \leq N_\alpha + N(D_\alpha)$. Applied to the generic path γ (of length bounded by $t(n)$), this gives $N_\gamma \leq 2^{(\log n)^{O(1)}} = o(2^n)$. This shows that $X_\gamma \cap \mathrm{KP}_n \neq \emptyset$ when n is large enough. That is absurd since the points of X_γ are rejected (because $\dim X_\gamma = n$). □

3 Reals with order

First we shall recall a fact : any language in $\mathrm{NP}_{\mathbb{R}_{ovs}}$ can be decided by a polynomial-time algorithm over $\mathbb{R}_{ovs}$ with the help of a boolean NP oracle [5]. This implies that any NP-complete problem $L \subseteq \{0,1\}^\infty \subseteq \mathbb{R}^\infty$ is $\mathrm{NP}_{\mathbb{R}_{ovs}}$-complete for Turing reduction. Of course such a language verifies $\dim(L \cap \mathbb{R}^n) = 0$, so there exist sparse NP-complete problems for $\leq_T$ over $\mathbb{R}_{ovs}$.

We conjecture that this is no longer true for the many-one reduction. Let us introduce a variant $\mathrm{KP}'_{\mathbb{R}}$ of the real knapsack problem. For a function t we define $\mathrm{KP}'_{\mathbb{R}}[t] \cap \mathbb{R}^{n+1}$ by

$$\{(y, x_1, \ldots, x_n) \in \mathrm{KP}_{\mathbb{R}}\ \forall \{a_1, \ldots, a_n\} \in \{-2^{t(n)}, \ldots, 2^{t(n)}\}^n \backslash \{0\}\ \sum_i a_i x_i \neq 0\}$$

Thus $\mathrm{KP}'_{\mathbb{R}}[t] \cap \mathbb{R}^{n+1}$ is the union of 2^n hyperplanes with some holes of dimension $n-1$. We have the following proposition.

Proposition 2 *If a language L such that $\dim(L \cap \mathbb{R}^n) = O(\log n)$ is $\mathrm{NP}_{\mathbb{R}_{ovs}}$-hard for $\leq_m$, then there is polynomial t and a real language A such that $\mathrm{KP}'_{\mathbb{R}}[t] \subseteq A \subseteq \mathrm{KP}_{\mathbb{R}}$ and $A \in \mathrm{P}_{\mathbb{R}_{ovs}}$.*

Proof of the proposition : we suppose there exists such a language L. Thus $\dim(L \cap \mathbb{R}^n) \leq a \log n$. As $\mathrm{KP}_{\mathbb{R}} \in \mathrm{NP}_{\mathbb{R}_{ovs}}$, we have a polynomial time reduction φ from $\mathrm{KP}_{\mathbb{R}}$ to L. Let t be a polynomial bounding the running time of φ. We now describe a polynomial time algorithm that decides $\mathrm{KP}'_{\mathbb{R}}[t]$. Let $x \in \mathbb{R}^n$. Its image by φ lies in $\mathbb{R}^{\leq t(n)}$. Let P_x be the set of point following the same path as x in φ. Performing a symbolic computation along this path allows us to obtain a system of affine equations describing P_x. Moreover, we can compute in polynomial time the dimension of P_x. If $\dim P_x \leq n-2$ we reject the input. If $\dim P_x = n-1$, we compute an equation of the affine closure g of P_x. If g is a hyperplane of $\mathrm{KP}_{\mathbb{R}}$ and $x \in g$ we accept; otherwise we reject. Now we suppose $\dim P_x = n$. The restriction of φ to P_x is an affine function. We note L_x its linear part : $L_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $m \leq bn^c$. The algorithm now searches the hyperplanes of $\mathrm{KP}_{\mathbb{R}}$ that intersect P_x with dimension $n-1$. We suppose now that there is such a hyperplane and note h its direction. By hypothesis, there is a constant a such that $\dim(L \cap \mathbb{R}^n) \leq a \log n$. First it is impossible that $\mathrm{rg} L_x > a\log(bn^c) + 1$. Indeed that would imply $\dim L_x(h) > a\log(bn^c)$ and $L_x(h)$ could not be completely accepted (because it would be included in a finite union of affine subspaces of dimension $a\log(bn^c)$). Thus $\mathrm{rg} L_x \leq a\log(bn^c) + 1$, that is to say $\dim \mathrm{Ker} L_x \geq n - a\log(bn^c) - 1$. Remark that $\mathrm{Ker} L_x \subseteq h$ otherwise all the point of P_x would be accepted. It remains to find all the hyperplanes of $\mathrm{KP}_{\mathbb{R}}$ whose direction contains $\mathrm{Ker} L_x$. This can be done by applying the algorithm suggested in the proof of Lemma 1. The algorithm accepts if the input x belongs to one of these at most $2^{n-(n-a\log(bn^c)-1)} = n^{O(1)}$ hyperplanes. $\square$

Remark 1 *We could also accept in the case when* $\dim P_x \leq n-2$, *deciding a language that would contain* $\mathrm{KP}_{\mathbb{R}}$.

We conjecture that, unless P = NP, there exists no t and A such that $\mathrm{KP}'_{\mathbb{R}}[t] \subseteq A \subseteq \mathrm{KP}_{\mathbb{R}}$ and $A \in \mathrm{P}_{\mathbb{R}_{ovs}}$. However, we cannot prove it even for $t(n) = 1$.

Acknowledgements

We would like to thank Felipe Cucker for encouraging us to look at these questions.

References

[1] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation.* Springer-Verlag, 1998.

[2] L. Blum, M. Shub, and S. Smale. On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. *Bulletin of the American Mathematical Society*, 21(1):1–46, July 1989.

[3] J. Cai and M. Ogihara. Sparse sets versus complexity classes. In L. Hemaspaandra and A. Selman, editors, *Complexity theory retrospective II.* Springer, 1997.

[4] F. Cucker, P. Koiran, and M. Matamala. Complexity and dimension. *Information Processing Letters*, 62:209–212, 1997.

[5] H. Fournier and P. Koiran. Lower bounds are not easier over the reals: Inside PH. LIP Research Report 99-21, Ecole Normale Supérieure de Lyon, 1999.

[6] B. Poizat. *Les Petits Cailloux.* Nur Al-Mantiq Wal-Ma'rifah **3**. Aléas, Lyon, 1995.