Laboratoire de l'Informatique du Parallélisme

Ecole Normale Supérieure de Lyon
Unité de recherche associée au CNRS n°1398

The Real Dimension Problem is $\mathrm{NP}_{\mathbb{R}}$*-Complete*

Pascal Koiran

Octobre 1997

Research Report N$^{\mathrm{O}}$ 97-36

Ecole Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : (+33) (0)4.72.72.80.00 Télécopieur : (+33) (0)4.72.72.80.80
Adresse électronique : lip@lip.ens−lyon.fr

The Real Dimension Problem is $NP_{\mathbb{R}}$-Complete

Pascal Koiran

Octobre 1997

Abstract

We show that computing the dimension of a semi-algebraic set of $\mathbb{R}^n$ is a $NP_{\mathbb{R}}$-complete problem in the Blum-Shub-Smale model of computation over the reals. Since this problem is easily seen to be $NP_{\mathbb{R}}$-hard, the main ingredient of the proof is a $NP_{\mathbb{R}}$ algorithm for computing the dimension.

Keywords: semi-algebraic sets, dimension, NP-completeness, Blum-Shub-Smale model.

Résumé

On montre que le calcul de la dimension d'un ensemble semi-algébrique de $\mathbb{R}^n$ est un problème $NP_{\mathbb{R}}$-complet dans le modèle de Blum-Shub-Smale de calcul sur les nombres réels. Puisqu'il est facile de voir que ce problème est $NP_{\mathbb{R}}$-dur, le principal ingrédient de la preuve est un algorithme $NP_{\mathbb{R}}$ de calcul de la dimension.

Mots-clés: ensembles semi-algébriques, dimension, NP-complétude, modèle de Blum-Shub-Smale.

The Real Dimension Problem is $\mathrm{NP}_{\mathbb{R}}$-Complete

Pascal Koiran
koiran@lip.ens-lyon.fr

October 7, 1997

Abstract

We show that computing the dimension of a semi-algebraic set of $\mathbb{R}^n$ is a $\mathrm{NP}_{\mathbb{R}}$-complete problem in the Blum-Shub-Smale model of computation over the reals. Since this problem is easily seen to be $\mathrm{NP}_{\mathbb{R}}$-hard, the main ingredient of the proof is a $\mathrm{NP}_{\mathbb{R}}$ algorithm for computing the dimension.

Keywords: semi-algebraic sets, dimension, NP-completeness, Blum-Shub-Smale model.

1 Introduction

This paper is a continuation of [14], which dealt with the dimension of complex algebraic varieties. Here we wish to compute the dimension of semi-algebraic sets. This can be formalized as a decision problem $\mathrm{DIM}_{\mathbb{R}}$. An instance of $\mathrm{DIM}_{\mathbb{R}}$ consists of a semi-algebraic $S \subseteq \mathbb{R}^n$ together with an integer $d \leq n$ (to be precise one should specify how S is represented, see section 1.1 for details). An instance is accepted if S has dimension at least d. We also consider for each fixed value of d the restriction $\mathrm{DIM}_{\mathbb{R}}^d$ of $\mathrm{DIM}_{\mathbb{R}}$. For instance, $\mathrm{DIM}_{\mathbb{R}}^0$ is the problem of deciding whether a semi-algebraic set has dimension ≥ 0, i.e., is nonempty.

This paper contributes to the still rather short list of $\mathrm{NP}_{\mathbb{R}}$-complete problems. The canonical $\mathrm{NP}_{\mathbb{R}}$-complete problem $4\mathrm{FEAS}_{\mathbb{R}}$ (feasibility of a polynomial equation of degree at most 4) was exhibited in [4]. A few other examples can be found in [9]. Here we show that $\mathrm{DIM}_{\mathbb{R}}$, and $\mathrm{DIM}_{\mathbb{R}}^d$ for any $d \geq 0$, are $\mathrm{NP}_{\mathbb{R}}$-complete problems. We emphasize that the situation is different than for most NP-complete combinatorial problems: as in [14], the dimension problem is easily seen to be NP-hard. It is the fact that $\mathrm{DIM}_{\mathbb{R}}$ *is* in $\mathrm{NP}_{\mathbb{R}}$ which is interesting. Thus this $\mathrm{NP}_{\mathbb{R}}$-completeness result should be viewed as a "positive" result. The technical tools are roughly the same as in the complex case ("generic quantifiers" and transcendence degree arguments). Some aspects of the proof are more involved than in [14], while others are actually simpler (see in particular the remark before (2) in section 3.1).

For polynomials with integer coefficients we are also interested in the classical (bit cost) complexity. We show that the corresponding problems (4FEAS and DIM) can be reduced to each other in polynomial time. Finally, the randomized and deterministic complexity of $\mathrm{DIM}_{\mathbb{R}}$ is touched upon in section 5.

1.1 Representation of semi-algebraic sets

Our results have very little dependence on the choice of a representation for semi-algebraic sets. It is customary to represent them as unions of basic semi-algebraic sets of the form

$$P_1(x)\ \Delta_1\ 0; \cdots ; P_m(x)\ \Delta_m\ 0 \tag{1}$$

with $\Delta_i \in \{>; \geq, =; \leq; <\}$. Since the dimension of a union is the maximum of the dimensions, one could without loss of generality work with basic semi-algebraic sets only.

The main theorem of this paper is the positive result that $\mathrm{DIM}_{\mathbb{R}}$ is in $\mathrm{NP}_{\mathbb{R}}$. It is thus desirable to work with a representation scheme for semi-algebraic sets which is as powerful as possible. Arithmetic circuits provide an appealing alternative to (1). In this case, S is represented by a circuit made of addition, multiplication and sign gates, which, on an input $x \in \mathbb{R}^n$, outputs 1 iff and only if $x \in S$. In fact, $\mathrm{NP}_{\mathbb{R}}$-completeness still holds for the even more powerful scheme in which S is represented by by an existential formula (this is also true over $\mathbb{C}$). For the sake of simplicity we will stick to (1) in the remainder of this paper, and use a sparse representation for the P_i's. As in [14], the $\mathrm{NP}_{\mathbb{R}}$-completeness result still holds for the dense representation and polynomials of degree at most 2 (here a single polynomial equation of degree at most 4 would suffice).

The defining formula for S will be denoted $\phi(x)$. If we wish to emphasize the dependence of ϕ on a tuple of parameters $a \in \mathbb{R}^p$, we will also write $\phi(a, x)$.

2 Background

The standard references for real algebraic geometry are [2] and [5].

2.1 Quantifier Elimination

We recall that the total degree σ of a first-order formula Φ is the sum of the polynomials appearing in Φ. It is convenient to always have $\sigma \geq 2$, so we will in fact define σ as $2 + \sum_{i=1}^{m} \deg p_i$, where $p_1, \ldots, p_m$ are the polynomials appearing in Φ.

This effective quantifier elimination result follows from the recent work on single-exponential algorithms in real geometry (in fact more precise bounds can be found in, e.g., [1] or [17]).

Theorem 1 *Let $\Phi(x)$ be a first-order formula with a total of n variables and $l \leq n$ free variables (thus $x \in R^l$). Assume that Φ is in prenex form with w blocks of quantifiers, has total degree σ, and that the polynomials in Φ have integer coefficients of bit length at most L. Let $n_1, \ldots, n_w$ be the lengths of the quantifier blocks (thus $n = l + \sum_{i=1}^{w} n_i$).*

If Φ is a closed formula ($l = 0$), its truth can be decided in time $\sigma^{2^{O(w)} \prod_k n_k}$ in the real number model.

For $l \geq 1$, $\Phi(x)$ is equivalent to a quantifier-free formula $\Psi(x)$ of the form:

$$\bigvee_{i=1}^{I} \bigwedge_{j=1}^{J_i} (Q_{ij}(x) \Delta_{ij} 0),$$

where Δ_{ij} is one of the 6 standard relations ($>, \geq, =, \neq, \leq, <$), $I = \sigma^{2^{O(w)} l \prod_k n_k}$, and J_i and the degrees of the polynomials Q_{ij} are bounded by $\sigma^{2^{O(w)} \prod_k n_k}$. These polynomials have integer coefficients of bit length at most $(L + l).\sigma^{2^{O(w)} \prod_k n_k}$. Moreover Ψ can be constructed in time $\sigma^{2^{O(w)} l \prod_k n_k}$ in the real number model.

2.2 Real Computation and Complexity

Here we will just recall the definition of $\mathrm{NP}_{\mathbb{R}}$ (see [3, 4, 16] for more information on the Blum-Shub-Smale model). A problem $A \subseteq \mathbb{R}^{\infty}$ is in $\mathrm{NP}_{\mathbb{R}}$ if there exists a problem $B \in \mathrm{P}_{\mathbb{R}}$ and a polynomial p such that for any $x \in \mathbb{R}^n$, $x \in A$ if there exists $y \in \mathbb{R}^{p(n)}$ such that $\langle x, y \rangle \in B$ (y is the "certificate" that $x \in A$).

This means essentially that for each n, $A \cap \mathbb{R}^n$ can be defined by an existential formula $F_n(x)$ of size polynomial in n (the free variable x lives in $\mathbb{R}^n$).

In order to recover the definition above, two conditions must be added:

(i) There exists a fixed tuple $a_1, \ldots, a_p$ of real numbers such that for every n the parameters of F_n are in $\{a_1, \ldots, a_p\}$ (so we will write $F_n(x, y)$ instead of $F_n(x)$; $A \cap \mathbb{R}^n$ is then defined by $F_n(x, a)$).

The $\mathrm{NP}_{\mathbb{R}}$ algorithms exhibited in this paper will be parameter-free. If one just adds condition (i), the class $\mathbb{NP}_{\mathbb{R}}$ defined by Poizat [16] is obtained (a short summary of this point of view can be found in [7]). For $\mathrm{NP}_{\mathbb{R}}$ there is an additional uniformity condition:

(ii) $F_n(x, y)$ can be produced in polynomial time by a (standard) Turing machine.

The main point here is the polynomial bound on the size of F_n. The uniformity condition may also lead to additional complications (this is certainly the case in this paper and in [14]). Over the reals, this condition is redundant if arbitrary real parameters are allowed (a family of circuits or formulas can be encoded in the digits of a real parameter), so that $\mathrm{P}_{\mathbb{R}} = \mathbb{P}_{\mathbb{R}}$ and $\mathbb{NP}_{\mathbb{R}} = \mathrm{NP}_{\mathbb{R}}$.

3 Generic Quantifiers

3.1 Efficient Elimination

We will use a non-standard quantifier $\exists^*$ which has the following meaning: if $F(v)$ is a first-order formula where the free variable v lives in $\mathbb{R}^d$, we say that $\mathbb{R} \models \exists^* v\ F(v)$ if the subset of $\mathbb{R}^d$ defined by F has nonempty interior. It is then natural to define another quantifier $\forall^*$ by: $\forall^* v\ F(v) \ \equiv\ \neg\exists^* v \neg F(v)$. That is, $\mathbb{R} \models \forall^* v\ F(v)$ if the set defined by F is dense in $\mathbb{R}^d$ (and in this case it contains an open dense set). Formulas involving generalized quantifiers will sometimes be called *generalized formulas* when there is a risk of confusion. Over $\mathbb{C}$ it is not completely obvious that generalized formulas can be replaced by ordinary first-order formulas in a "concise" manner (see [14] or better [13]). In the real case this is of course no problem since $\exists^* v\ F(v)$ is equivalent to

$$\exists x \in \mathbb{R}^d\ \exists \epsilon > 0\ \forall y \in \mathbb{R}^d\ [||x-y||^2 \le \epsilon \Rightarrow F(y)] \tag{2}$$

However this transformation is not quite satisfactory because (2) has two quantifier blocks. It will be seen shortly that one can do better. We begin with a series of simple lemmas.

Lemma 1 *Let $G(v)$ be a quantifier-free first-order formula where the free variable v lives in $\mathbb{R}^d$. Let $p_1, \ldots, p_m$ be the polynomials appearing in G. If there exists an $x \in \mathbb{R}^d$ satisfying G such that $p_i(x) \neq 0$ for $i = 1, \ldots, m$ then $\mathbb{R} \models \exists^* v\ G(v)$.*

Proof. The sign of the p_i's remain constant in a neighbourhood of x. Since the satisfaction of G depends only on those signs all points in the neighbourhood satisfy G. $\Box$

Proposition 1 *Let $F(v)$ a first-order formula where the free variable v lives in $\mathbb{R}^d$, and $K \subseteq \mathbb{R}$ the field generated by the parameters of F. Then $\mathbb{R} \models \forall^* v\ F(v)$ iff and only if for any $a = (a_1, \ldots, a_d)$ of transcendence degree d over K, $\mathbb{R} \models F(a)$.*

Proof. Since quantifier elimination does not require any introduction of new parameters, we will assume that F is quantifier free. If $\mathbb{R} \models F(a)$ for an a with transcendence degree d, the conclusion follows from Lemma 1 applied to $G = \neg F$. The converse holds because $\mathbb{R}$ has infinite transcendence degree. $\Box$

Lemma 2 *Let K be a subfield of $\mathbb{R}$ and $a = (a_1, \ldots, a_k)$ a sequence of elements of $\mathbb{R}$ that are algebraically independent over K. For any $s < k$ and $(v_1, \ldots, v_s) \in \mathbb{R}^s$, there exists a subsequence $(a_{i_j})_{1 \le j \le k-s}$ whose elements are algebraically independent over the the field $K' = K(v_1, \ldots, v_s)$.*

Proof. Let K'' be the field extension of K' generated by the a_i's: $\text{tr.deg}_{K'}K'' \geq k - s$ since $\text{tr.deg}_K K'' = \text{tr.deg}_{K'}K'' + \text{tr.deg}_K K'$ (this is e.g. the corollary of Theorem 4 in section V.14.3 of [6]), $\text{tr.deg}_K K' \leq s$ and $\text{tr.deg}_K K'' \geq k$ by definition of a. Let B be a transcendence base of K'' over K' made up of elements of a. B has at least $k - s$ elements, and they are algebraically independent over K' as needed. $\square$

Lemma 3 *Let K be a subfield of $\mathbb{R}$, $x \in \mathbb{R}^d$ and $\epsilon \in \mathbb{R}$, $\epsilon \neq 0$. If the components of $y \in \mathbb{R}^d$ are algebraically independent over the field $K(x, \epsilon)$ then the components of $x + \epsilon y$ are algebraically independent over K.*

Proof. We need to show that for $P \in K[X_1, \ldots, X_d]$, if $P(x + \epsilon y) = 0$ then P is identically 0. $P(x + \epsilon X)$ can be written as a polynomial $P_{x,\epsilon}(X)$ with coefficients in $K[x, \epsilon]$. If $P(x + \epsilon y) = 0$ then $P_{x,\epsilon}(y) = 0$, hence $P_{x,\epsilon}$ is identically 0 by the hypothesis on y. Thus $P(x + \epsilon a) = 0$ for any $a \in \mathbb{R}^d$. We conclude that $P \equiv 0$ since $\epsilon \neq 0$. $\square$

Let $F(u, v)$ be a first-order formula where $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^d$ (one can think of u as a "parameter" and v as an "instance"). Let $\tilde{F}(u, y_1, \ldots, y_{d+p+2})$ be the formula:

$$\exists x \in \mathbb{R}^d \; \exists \epsilon > 0 \; \bigwedge_{i=1}^{d+p+2} F(u, x + \epsilon y_i).$$

Here each variable y_i is in $\mathbb{R}^d$. Then $W(F)$ denotes the set of sequences $y = (y_1, \ldots, y_{d+p+2}) \in \mathbb{R}^{d(d+p+2)}$ such that

$$\forall u \in \mathbb{R}^p \; [\tilde{F}(u, y_1, \ldots, y_{d+p+2}) \Leftrightarrow \exists^* v \; F(u, v)]. \tag{3}$$

Theorem 2 *For any first-order formula F, $W(F)$ is dense in $\mathbb{R}^{d(d+p+2)}$.*

Proof. Let K be the subfield of $\mathbb{R}$ generated by the parameters of F. By Proposition 1, it suffices to show that $y \in W(F)$ whenever the components of y are algebraically independent over K.

Fix any $u \in \mathbb{R}^p$. If $\mathbb{R} \models \exists^* v \; F(u, v)$ it is clear that $\mathbb{R} \models \tilde{F}(u, y)$ for every $y \in \mathbb{R}^{d(d+p+2)}$. We now examine the case $\mathbb{R} \models \forall^* v \; \neg F(u, v)$. Take $y = (y_1, \ldots, y_{d+p+2})$ with coordinates that are algebraically independent over K, and fix any $x \in \mathbb{R}^d$ and $\epsilon > 0$. By Lemma 2, at least $d(d + p + 2) - (d + p + 1)$ among the $d(d + p + 2)$ components of the y_i's are algebraically independent over $K(u, x, \epsilon)$. Thus there exists at least one y_i with coordinates that are algebraically independent over $K(u, x, \epsilon)$. By Lemma 3 the coordinates of $x + \epsilon y_i$ are then algebraically independent over $K(u)$. Thus $\mathbb{R} \models \neg F(u, x + \epsilon y_i)$ by Proposition 1, and therefore $\mathbb{R} \models \neg \tilde{F}(u, y)$. $\square$

As we shall see in Section 3.2, the density of $W(F)$ implies that one can deterministically construct a point in this set (or just choose one at random). Thus Theorem 2 makes it possible to replace a generic quantifier by an existential formula.

When there are no parameters ($p = 0$) the sequences in $W(F)$ have length $d+2$. The example of the unit sphere ($F(v) \equiv [v_1^2 + \cdots + v_d^2 = 1]$) shows that this bound cannot be improved in general (this follows from the fact that generically, $d+1$ points in $\mathbb{R}^d$ lie on the same $(d-1)$-sphere).

3.2 Explicit Construction

Lemma 4 *Let $G(v)$ be a quantifier-free formula such that $\mathbb{R} \models \forall^* v \in \mathbb{R}^d\ G(v)$. Assume that the polynomials in G are of degree at most D, with integer coefficients bounded by M in absolute value. Any point $\alpha = (\alpha_1, \ldots, \alpha_d)$ satisfying $\alpha_1 \geq M+1$ and $\alpha_j \geq 1 + M(D+1)^{j-1}\alpha_{j-1}^D$ for $j \geq 2$ satisfies G.*

Proof. Let $p_1, \ldots, p_m$ be the polynomials occurring in G. Then α satisfies $p_i(\alpha) \neq 0$ for any $i = 1, \ldots, m$. A proof of this simple fact can be found in Lemma 2 of [12] (here we have a corrected a mistake in the statement of that lemma). Hence α satisfies G by Lemma 1. $\square$

Note that the sequence in this lemma can be constructed in a polynomial number of arithmetic operations (more precisely in $O(\log\log M + d\log D)$ operations starting from the integer 1). Nonetheless the components of α are of bit size exponential in d.

Lemma 4 can be applied to a quantified formula if we eliminate quantifiers first.

Corollary 1 *Let G be a prenex formula such that $\mathbb{R} \models \forall^* v \in \mathbb{R}^d\ G(v)$. Let σ be its total degree, w the number of quantifier blocks, and n the total number of variables. If the parameters in G are integers of bit size at most L, one can construct in $O(\log L) + O(n)^w \log\sigma$ arithmetic operations an integer point that satisfies G. This point depends only on L, n and σ.*

Proof. Immediate from Theorem 1 and Proposition 4. $\square$

We are now ready to give an explicit construction of a point in $W(F)$.

Theorem 3 *Let $F(u, v)$ be a prenex formula with a total number of n variables, w quantifier blocks, and m atomic predicates of degree at most D with integer coefficients of bit size at most L. One can construct in $O(\log L) + n^{O(w)} \log(mD)$ arithmetic operations an integer point in $W(F)$.*

Proof. For the sake of clarity, we consider quantifier-free formulas first. Recall that $W(F)$ is defined by (3). This formula can be transformed into an "ordinary" first-order formula if we substitute (2) to the generic quantifier in (3). (This transformation is not so easy in the complex case.) When put in prenex form, the resulting formula has $O(n^2)$ variables and $O(1)$ quantifier blocks. It involves $O(mn)$ atomic predicates atomic predicates of degree at most $2D$ with coefficients of bit size at most $L+D$. The result then follows from Corollary 1 since we know from Theorem 2 that $W(F)$ is dense.

In the general case, we can first eliminate quantifiers in F with Theorem 1. $\square$

4 $\mathbf{NP_{\mathbb{R}}}$-Completeness

We will show as an intermediate result that the "projection problem" $\mathrm{PROJ}_{\mathbb{R}}$ is $\mathrm{NP}_{\mathbb{R}}$-complete. An instance of this problem consists of a semi-algebraic $S \subseteq \mathbb{R}^n$ together with an integer $d \leq n$. An instance is positive if the image of S by the projection $\pi_d : \mathbb{R}^n \rightarrow \mathbb{R}^d$ on the first d coordinates has a non-empty interior.

Theorem 4 $\mathrm{PROJ}_{\mathbb{R}}$ *is* $\mathrm{NP}_{\mathbb{R}}$*-complete.*

Proof. The projection $\pi_d(S)$ is defined by a formula $F(u, x)$:

$$\exists z \in \mathbb{R}^{n-d} \phi(u, x, z)$$

where the free variable x is in $\mathbb{R}^d$. Here $u \in \mathbb{R}^p$ is the tuple of nonzero parameters occurring in ϕ (so that $\phi(.,.,.)$ is parameter-free). By definition of $W(F)$, $\pi_d(S)$ has nonempty interior if $\mathbb{R} \models \tilde{F}(y_1, \ldots, y_{d+p+2})$ where $(y_1, \ldots, y_{d+p+2})$ is any sequence in $W(F)$. By Theorem 3 such a sequence can be constructed in polynomial time. Moreover, $\tilde{F}$ can be written in prenex form as an existential formula of polynomial size since F itself is existential (there are $(d+p+2)(n-d)+d+1$ quantified variables in the resulting formula). This shows that $\mathrm{PROJ}_{\mathbb{R}} \in \mathrm{NP}_{\mathbb{R}}$.

Its $\mathrm{NP}_{\mathbb{R}}$-hardness follows from a (trivial) reduction of $4\mathrm{FEAS}_{\mathbb{R}}$ to $\mathrm{PROJ}_{\mathbb{R}}$: a polynomial $p \in \mathbb{R}[X_2, \ldots, X_{n+1}]$ has a real root if and only if the projection on the first coordinate x_1 of the set $S = \{x \in \mathbb{R}^{n+1};\ p(x_2, \ldots, x_{n+1}) = 0\}$ has a nonempty interior. $\square$

Theorem 5 $\mathrm{DIM}_{\mathbb{R}}$ *and, for any* $d \geq 0$*,* $\mathrm{DIM}^d_{\mathbb{R}}$ *are* $\mathrm{NP}_{\mathbb{R}}$*-complete problems.*

Proof. A semi-algebraic set S has dimension at least d if there exists a d-dimensional coordinate subspace on which S has a projection with a nonempty interior. Hence $\mathrm{DIM}_{\mathbb{R}}$ can be solved by the following $\mathrm{NP}_{\mathbb{R}}$ algorithm: guess d distinct indices $i_1, \ldots, i_d$ in $\{1, \ldots, n\}$ and (renumbering variables if necessary) decide with the $\mathrm{NP}_{\mathbb{R}}$ algorithm of Theorem 4 whether the projection of S on the corresponding coordinate subspace has nonempty interior.

One can show as in the proof of Theorem 4 that $\mathrm{DIM}^d_{\mathbb{R}}$ (and *a fortiori* $\mathrm{DIM}_{\mathbb{R}}$) are $\mathrm{NP}_{\mathbb{R}}$-hard (just add d dummy variables to a polynomial equation). $\square$

For systems with integer coefficients in the bit model of computation, there is currently no hope of proving a completeness result since even the exact complexity of 4FEAS is unknown (in terms of structural complexity, this problem is only known to lie somewhere between NP and PSPACE). However, one can show that DIM and 4FEAS are reducible to each other in polynomial time.

Theorem 6 DIM *is polynomially equivalent to* 4FEAS.

Proof Sketch. The reduction of $4\mathrm{FEAS}_{\mathbb{R}}$ to $\mathrm{DIM}_{\mathbb{R}}$ provides a reduction of 4FEAS to DIM.

The proof that $\mathrm{DIM}_{\mathbb{R}}$ is in $\mathrm{NP}_{\mathbb{R}}$ provides a reduction of $\mathrm{DIM}_{\mathbb{R}}$ to $4\mathrm{FEAS}_{\mathbb{R}}$. In the bit model of computation this yields a reduction of DIM to 4FEAS (note that one can take $p = 0$ in this case). Unfortunately this reduction does not work in polynomial time since it entails the computation of integer points with exponential bit size. Instead of computing the α_j's of Lemma 4, we can introduce new variables to represent them. The corresponding reduction is polynomial-time as needed. $\square$

5 Randomized and Deterministic Algorithms

In this section we wish to take a closer look at the complexity of sequential algorithms for $\mathrm{DIM}_{\mathbb{R}}$. As in the $\mathrm{NP}_{\mathbb{R}}$-completeness proof, we reduce $\mathrm{DIM}_{\mathbb{R}}$ to $\mathrm{PROJ}_{\mathbb{R}}$. The resolution of this auxiliary problem is by far the most expensive step.

5.1 Reduction to $\mathrm{PROJ}_{\mathbb{R}}$

We say that a semi-algebraic set S of dimension $\geq d$ is in *normal position* with respect to a subset of d distinct variables $\{X_{i_1}, \ldots, X_{i_d}\}$ if the projection of S on the corresponding d-dimensional coordinate subspace has nonempty interior. The proof of Theorem 5 suggests to enumerate all such subsets, and for each one to check whether S is in normal position. This can be done without affecting the overall complexity bound (see section 5.2), but there is a more practical solution: performing a sufficiently "generic" linear transformation on S will put this set in normal position with respect to the first d variables. Unfortunately, such a transformation can blow up the system's size. In the complex case there is a way around this difficulty: a definable set has dimension at least d if it has a nonempty intersection with a "generic" affine subspace of dimension $n - d$. A similar property holds over the reals: as in the complex case [14], we can just *pretend* to perform a linear transformation. That is, we consider the variety $\hat{S} \subseteq \mathbb{R}^{2n}$ defined by the system

$$\begin{cases} \phi(x) \\ y = Ax. \end{cases} \tag{4}$$

where A is the matrix of the linear transformation. We recall that ϕ is a system of m (in)equations defining S. It is clear that $\pi_d(AS) = \hat{\pi}_d(\hat{S})$ where $\hat{\pi}_d : \mathbb{R}^{2n} \to \mathbb{R}^d$ denotes projection on the variables $y_1, \ldots, y_d$. Note that the last $n-d$ equations can be dropped from this system since they are automatically satisfied (from the relation $y = Ax$) if a solution exists for $x_1, \ldots, x_n$ and $y_1, \ldots, y_d$. Therefore we have to solve an instance of $\text{PROJ}_{\mathbb{R}}$ made of $m+d$ inequations in $n+d$ variables (here we are PROJecting on the variables $y_1, \ldots, y_d$). These observations can be summarized by the following principle (which does not use the hypothesis that S is semi-algebraic in any essential way).

A semi-algebraic set S has dimension at least d if given a generic linear subspace $Bx = 0$ of dimension $n-d$, the affine subspace $y = Bx$ has a nonempty intersection with S for y in a subset of $\mathbb{R}^d$ with nonempty interior.

In a randomized implementation, the coefficients of B would be randomly drawn integers. It is possible to work out a polynomial bound on their bit size. We will not go into the details since they are essentially the same as in the complex case. It is also possible to construct a suitable B deterministically, see again [14].

5.2 Complexity of $\text{PROJ}_{\mathbb{R}}$

It is almost a folklore result that $\text{PROJ}_{\mathbb{R}}$ (and thus $\text{DIM}_{\mathbb{R}}$) can be solved in time $(sD)^{O(n^2)}$ by quantifier elimination. Since there does not seem to be an appropriate reference in the literature, we sketch the proof below. As a first attempt, one can use (2) to express the fact that the projection of S has a nonempty interior. The resulting formula has 3 quantifier blocks since F is an existential formula in this case. It can therefore be decided in time $(sD)^{O(n^3)}$ with the algorithms of [1] or [17]. To do better, one computes in time $(sD)^{O(n^2)}$ with the algorithm of Theorem 1 a quantifier-free formula $\Psi(x)$ defining $\pi_d(S)$. This formula is a disjunction of $(sD)^{O(n^2)}$ conjunctions. $\pi_d(S)$ has nonempty interior if one of the conjunctions defines a set with nonempty interior. Consider a conjunction C of constraints of the form $p_i \; \Delta_i \; 0$ where p_i is a non-constant polynomial and Δ_i is a standard relation. The set defined by C has nonempty interior if no Δ_i is an equality and if the formula C' obtained from C by replacing every large inequality by a strict inequality is satisfiable. The satisfiability of C' can be decided in time $(sD)^{O(n^2)}$ since the p_i's are bounded in degree and number by $(sD)^{O(n)}$. This is also an upper bound on the overall running time of the algorithm.

Theorem 4 also yields a $(sD)^{O(n^2)}$ algorithm since it reduces $\mathrm{PROJ}_{\mathbb{R}}$ to the satisfaction of an existential formula in $O(n^2)$ variables. In practice on would not perform a deterministic reduction as in the proof of that theorem. Instead a sequence in $W(F)$ would be drawn at random. To see how a bit size bound can be worked out, we refer again the interested reader to [14].

6 Final Remarks

The main open problem is whether $\mathrm{DIM}_{\mathbb{R}}$ can be solved in time $(sD)^{O(n)}$. Some progress in this direction has been made in [18] where this bound is achieved for *smooth* semi-algebraic sets. In the complex case it is known that the dimension can always be computed within that time bound (and in fact in time $s^{O(1)}D^{O(n)}$). For instance this follows from the fact that the randomized reduction in [14] produces existential formulas with only $O(n)$ variables (see also [8, 10, 11, 15]). It is by no means clear whether a similarly "parsimonious" reduction exists in the real case. If this question turns out to have a positive answer, a $(sD)^{O(n)}$ bound for $\mathrm{DIM}_{\mathbb{R}}$ can be expected.

On the other hand, as we have already pointed out in section 3.1, life is sometimes easier over the reals than over the complex numbers. Consider for instance the problem of determining whether a complex algebraic variety has isolated points (this question is motivated by the problem of computing the dimensions of all components of a variety as in [10]; see also [11]). It is not clear whether this problem in in $\mathrm{PH}_{\mathbb{C}}$, the polynomial hierarchy over $\mathbb{C}$ (this amounts basically to asking whether the existence of isolated points is a property that can be expressed by first-order formulas of polynomial size with a bounded number of quantifier alternations). However, it is quite obvious that the corresponding problem over the reals is in $\mathrm{PH}_{\mathbb{R}}$.

Acknowledgments

I thank Marie-Françoise Roy for useful discussions on the deterministic complexity of the real dimension problem.

References

[1] S. Basu, R. Pollack, and M.-F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. *Journal of the ACM*, 43(6):1002–1045, 1996.

[2] R. Benedetti and J.-J. Risler. *Real algebraic and semi-algebraic sets.* Hermann, Paris, 1990.

[3] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation.* Springer Verlag, to appear.

[4] 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.

[5] J. Bochnak, M. Coste, and M.-F. Roy. *Géométrie Algébrique Réelle.* Springer-Verlag, 1987.

[6] N. Bourbaki. *Algèbre (Chapitres 4 à 7).* Masson, Paris, 1981.

[7] O. Chapuis and P. Koiran. Saturation and stability in the theory of computation over the reals. Technical Report 1997/3, Institut Girard Desargues, Université Claude Bernard Lyon I, 1997.

[8] A. Chistov. Polynomial-time computation of the dimension of algebraic varieties in zero-characteristic. *Journal of Symbolic Computation*, 22:1–25, 1996.

[9] F. Cucker and F. Roselló. On the complexity of some problems for the Blum, Shub & Smale model. In *Proceedings of Latin'92*, volume 583 of *Lecture Notes in Computer Science*, pages 117–129. Springer-Verlag, 1992.

[10] M. Giusti and J. Heintz. Algorithmes – disons rapides – pour la décomposition d'une variété algébrique en composantes irréductibles et équidimensionnelles. In T. Mora and C. Traverso, editors, *Effective Methods in Algebraic Geometry (MEGA'90)*, Progress in Mathematics 94, pages 169–194. Birkhäuser, 1991.

[11] M. Giusti and J. Heintz. La détermination des points isolés et la dimension d'une variété algébrique peut se faire en temps polynomial. In *Computational Algebraic Geometry and Commutative Algebra (Cortona, 1991)*, pages 216–256. Sympos. Math. XXXIV, Cambridge University Press, 1993.

[12] P. Koiran. Elimination of constants from machines over algebraically closed fields. *Journal of Complexity*, 13(1):65–82, 1997. Erratum on *http://www.ens-lyon.fr/~koiran.*

[13] P. Koiran. Elimination of parameters in the polynomial hierarchy. LIP Research Report 97-37, Ecole Normale Supérieure de Lyon, 1997.

[14] P. Koiran. Randomized and deterministic algorithms for the dimension of algebraic varieties. In *Proc. 38th IEEE Symposium on Foundations of Computer Science*, 1997.

[15] G. Matera and J. Torres. The space complexity of elimination theory: Upper bounds. In F. Cucker and M. Shub, editors, *Foundations of Computational Mathematics (Selected Papers of a Conference Held at IMPA in Rio de Janeiro)*, pages 267–276, 1997.

[16] B. Poizat. *Les Petits Cailloux*. Aléas, 1995.

[17] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. parts I, II, III. *Journal of Symbolic Computation*, 13(3):255–352, March 1992.

[18] N. Vorobjov. Computing dimensions of semi-algebraic sets. preprint, 1997.